

**Rapport scientifique et technique  
du projet RefDivinités  
d'interopérabilité de référentiels  
sur les divinités et héros du monde  
méditerranéen antique**

Projet lauréat CollEx-Persée 2019

Sébastien Dalmon, François Haas et Alyx Taounza-Jeminet

## Table des matières

Introduction.....	4
1. Identification des entités.....	6
1.1 Travail comparatif.....	6
1.1.1 Extraction des corpus.....	6
1.1.1.1 Extraction des données d'IdRef.....	6
1.1.1.2 Extraction des données de PACTOLS.....	8
1.1.2 Comparaison dans OpenRefine.....	11
1.2 Création des notices manquantes.....	11
2. Réflexion méthodologique avec les spécialistes.....	12
2.1 Familiariser le comité scientifique avec les enjeux.....	12
2.2 Évaluation des problématiques avec les spécialistes.....	13
2.2.1 Fixer le corpus.....	13
2.2.2 Redéfinition des terminologies.....	13
2.2.2.1 Terminologies des concepts.....	13
2.2.2.2 Terminologies des aires.....	14
2.3 L'établissement des sources.....	14
2.4 Maintenir le dialogue.....	15
3. Traitement des entités dans IdRef.....	16
3.1 Traitement pratique des notices.....	16
3.2 Normes fixées.....	17
3.2.1 Normes de translittération.....	18
3.2.2 Prendre en compte les évolutions du guide méthodologique et leur impact potentiel.....	19
3.2.3 Suivi des modifications.....	20
3.3 Signalement des problématiques à la BnF/Rameau et à l'Abes.....	20
4. Alignement des données.....	21
4.1 Quels alignements pour quels enjeux ?.....	21
4.1.1 BnF.....	21
4.1.2 VIAF.....	21
4.1.3 Wikidata.....	22
4.1.4 Library of Congress.....	22
4.1.5 Thesaurus du Getty, AAT et IA.....	22
4.2. Traiter l'alignement des données.....	23
4.2.1 Traitement manuel dans IdRef.....	23
4.2.2 Traitement automatisé dans IdRef.....	23
4.2.3 Réaliser un alignement dans PACTOLS.....	24
4.2.4 Traitement automatisé dans PACTOLS.....	24
4.2.5 Traitement automatisé vers Wikidata.....	25
4.3 Phases pratiques de traitement durant le projet RefDivinités.....	26
5. Valorisations et perspectives autour du projet.....	27

5.1 La transition vers PACTOLS 2 : des problématiques spécifiques .....	27
5.1.1 Propositions pour DARIAH .....	27
5.1.2 Préparatifs et projections pour le déplacement de branches.....	28
5.2 Chantiers de réindexation.....	28
5.3 Quel statut pour ces notices s'il ne s'agit plus de personnes avec la transition bibliographique ? .....	29
5.4 Enrichissements complémentaires : bases de données fermées et formes d'entités particulières .....	30
5.5 Communications autour du projet .....	31
Conclusion .....	32
Quelques chiffres :.....	32
Annexes .....	33
Annexe 1 : Liste des membres du conseil scientifique.....	33
Annexe 2 : Diagramme des « aires religieuses ».....	34
Annexe 3 : Arborescences et terminologies dans PACTOLS et IdRef en avril 2019.....	36
IdRef.....	36
PACTOLS .....	38

# Introduction

Le projet RefDivinités d'interopérabilité de référentiels sur les divinités et héros du monde méditerranéen antique, déposé en septembre 2018, a été sélectionné en janvier 2019 lors de la deuxième vague d'appels à projets de Collex-Persée pour le développement de la numérisation et des services à la recherche.

Il réunit la Bibliothèque interuniversitaire de la Sorbonne, délégataire en sciences de l'Antiquité au sein du réseau CollEx, et le réseau FRANTIQ, dans l'objectif de rapprocher et enrichir leurs outils d'indexation spécifiques. Au sein du SUDOC, qui réunit les collections bibliographiques universitaires, la BIS utilise les notices d'autorité du référentiel IdRef pour normaliser et décrire celles-ci. La gestion des notices tant bibliographiques que d'autorité est supervisée par l'Abes, mais leur création et enrichissement sont le fruit du travail de catalogueur.ses au sein des bibliothèques du réseau grâce au logiciel WinIBW. Le réseau FRANTIQ, Fédération et Ressources sur l'Antiquité, possède son propre catalogue commun indexé (CCI) pour l'interrogation des collections d'institutions spécialisés dans l'Antiquité, alimenté par les laboratoires, musées et bibliothèques fédérés. La description de ces ressources est possible grâce au thésaurus collaboratif PACTOLS qui gère les entités. Créé dans les années 1980, il est aujourd'hui développé et repensé grâce au logiciel OpenTheso.

L'objectif de ce projet est de pouvoir offrir un basculement simplifié d'un service à l'autre grâce à l'alignement des notices d'autorité d'IdRef et des entités de PACTOLS. Il permet aussi de réviser et enrichir le corpus choisi pour le projet. Celui-ci a été circonscrit aux "divinités et héros du monde méditerranéen antique". Il englobe les puissances et figures mythologiques allant des Gaules à la Perse sur une période qui va de la fin du IV<sup>e</sup> millénaire avant notre ère au V<sup>e</sup> siècle de notre ère. Le projet s'est limité aux religions polythéistes, excluant les personnages bibliques. Il permet enfin de repenser certaines logiques de classifications afin d'être en phase avec l'actualité de la recherche.

Dans IdRef, les notices d'autorité qui nous intéressent sont qualifiées de trois manières :

- "divinité(s) + origine" (ex. : Isis (divinité égyptienne) ;
- "mythologie + origine" (ex. : Ulysse (mythologie grecque) ;
- "personnage légendaire" (ex. : Didon (personnage légendaire).

Les personnages légendaires peuvent avoir des termes liés mais ne sont généralement pas dépendants d'un parent, contrairement aux puissances divines et figures mythologiques : les divinités ont pour terme générique "Dieux + origine" ou "Déesses + origine" (notons ici que le terme générique "Déesses" est lui-même assujéti à "Dieux"). "Dieux" est un terme spécifique à la fois de "Religion + origine" et de "Mythologie + origine". "Religion + origine" remonte à "Religions" ou "Religion + qualificatif de lieu". "Mythologie + origine" remonte à "Civilisation + qualificatif de lieu". L'ensemble peut connaître des variations ainsi que des intermédiaires culturels plus larges et discutables (celtique emboîte gaulois, classique emboîte grec et romain par exemple). Surtout, il faut signaler que dans IdRef, les relations ne sont pas automatiquement réciproques bien qu'entre notices d'autorité cette pratique soit encouragée.

Dans PACTOLS, la lecture est beaucoup plus linéaire, les relations automatiquement réciproques, et les entités se doivent d'appartenir à une arborescence, ne pouvant rester "orphelines". PACTOLS est un thésaurus normé selon le standard ISO 25964. Il était subdivisé en micro-thésaurus : Peuples / Anthroponymes / Chronologie / Toponymes / Œuvres / Lieux et sites archéologiques / Sujets (termes dont les initiales forment l'acronyme PACTOLS). Les

notices d'autorité qui nous intéressent étaient intégrées dans le micro-thésaurus "Anthroponymes", à l'intérieur de deux sous-groupes, "Divinités" et "Héros". Chacun de ces sous-groupes se subdivise ensuite en aires civilisationnelles : sous "Divinités", on trouve ainsi les catégories "Divinités égyptiennes", "Divinités grecques", "Divinités romaines", etc ; de la même manière, sous "Héros", on trouve les catégories "Héros grecs", "Héros orientaux", "Héros romains"... Les entités nommées se classent à l'intérieur de ces sous-groupes (avec parfois des sous-catégories intermédiaires, comme "Olympiens" sous "Divinités"). Par ailleurs, dans PACTOLS, le micro-thésaurus "Sujets" comprenait des concepts comme "religion", sous lequel on trouvait "divinités", qui se subdivisait lui-même en d'autres concepts : "divinités de lieux", "divinités fonctionnelles", "panthéon égyptien", "panthéon étrusque", etc... On distingue donc *divinités* et *panthéons* en tant que concepts des *Divinités* en tant que catégorie regroupant un ensemble de puissances. Cette organisation logique d'origine présente, malgré tout, des contresens (difficile de considérer des noms d'êtres et créatures fantastiques comme des anthroponymes), dérives et redondances. Elle est bouleversée par la transition vers la deuxième version de PACTOLS sur les bases du Backbone Thesaurus de DARIAH qui devrait être achevée en 2022 et au sein duquel s'inscrit RefDivinités.

Un comité scientifique composé de 20 chercheuses et chercheurs spécialistes des religions anciennes a été mis en place. Le but de cette instance était d'obtenir une validation scientifique, pour les notices d'autorités concernées par le projet, qui soit au plus près des résultats de la recherche la plus actuelle. Les membres de ce comité, issus de douze établissements et rattachés à différents centres de recherche, présentaient des profils variés : enseignants-chercheurs en poste, émérites, plus ou moins au fait des humanités numériques. Trois réunions ont été organisées entre mai et octobre 2019, mais les contacts par mél, téléphone ou visioconférence ont également été largement utilisés, notamment avec la responsable du programme ERC « Mapping Ancient Polytheisms ». Un dialogue très fructueux s'est ainsi noué entre bibliothécaires et chercheurs, dans un esprit d'enrichissement mutuel. Les avis du comité scientifique ont permis de réfléchir à la fois sur le vocabulaire structurant et les modes de classification et distinction, en dialogue avec nos éclairages sur les logiques et standards de constitution d'un thésaurus et des usages des notices d'autorités. Cette collaboration a permis de faire émerger des propositions concrètes, afin de créer, de revoir et d'enrichir nombre de notices.

# 1. Identification des entités

## 1.1 Travail comparatif

Il s'agit d'abord de repérer quelles entités existent dans un référentiel et non dans l'autre, mais aussi de voir comment se distinguent certaines entités par leurs classification, variantes, etc. d'un référentiel à l'autre.

### 1.1.1 Extraction des corpus

#### 1.1.1.1 Extraction des données d'IdRef

Pour extraire les données d'IdRef, il existe la plateforme data.idref.fr. Celle-ci, par l'intermédiaire de son API Yasgui permet de réaliser une requête SPARQL extrêmement complète qui pourra prendre en compte une multiplicité de facteurs de recherche :

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX marcel: <http://id.loc.gov/vocabulary/relators/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdau: <http://rdaregistry.info/Elements/u/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>

select distinct ?id (GROUP_CONCAT(DISTINCT ?nom; SEPARATOR=' ; ') AS ?Entity)
(GROUP_CONCAT(DISTINCT ?altname; SEPARATOR=' ; ') AS ?Nomsalternatifs)
(GROUP_CONCAT(DISTINCT ?generique;
SEPARATOR=' ; ') AS ?arborescence) (GROUP_CONCAT(DISTINCT ?alignment;
SEPARATOR=' ; ') AS ?alignments)
(GROUP_CONCAT(DISTINCT ?source; SEPARATOR=' ; ') AS ?sources)
(GROUP_CONCAT(DISTINCT ?def; SEPARATOR='
; ') AS ?definition)

where {
{
?doc foaf:primaryTopic ?id.
```

```

?id skos:prefLabel ?nom.

filter                                     regex(?nom,
'([divinité])(mythologie)(mythologique|))([dieu])([déesse])(légendaire)')

OPTIONAL{?id skos:altLabel ?altname.}

OPTIONAL{?id skos:note|rdau:P60492 ?def.} OPTIONAL{?id skos:broader+
?idparent.

?idparent skos:prefLabel ?generique.}

OPTIONAL{?doc skos:editorialNote ?source.}

OPTIONAL {?id owl:sameAs ?alignment.}

}

UNION

{ ?doc foaf:primaryTopic ?id.

?id skos:altLabel ?altname.

filter                                     regex(?altname,
'([divinité])(mythologie)(mythologique|))([dieu])([déesse])(légendaire)')

OPTIONAL{?id skos:prefLabel ?nom.} OPTIONAL{?id skos:note|rdau:P60492 ?def.}
OPTIONAL{?id skos:broader+ ?idparent.

?idparent skos:prefLabel ?generique.}

OPTIONAL{?doc skos:editorialNote ?source.}

OPTIONAL {?id owl:sameAs ?alignment.}

}

UNION

{

?idgen skos:narrower ?id; skos:prefLabel ?generique. filter regex(?generique,
'([divinité])(mythologie)(mythologique|))([dieu])([déesse])(légendaire)|(Dieux)|(Déesse)
|(Pantheon)')

OPTIONAL{?id skos:prefLabel ?nom.}

OPTIONAL{?id skos:altLabel ?altname.} OPTIONAL{?id bibo:Document ?doc. ?doc
skos:note|rdau:P60492 ?def.}

OPTIONAL {?id owl:sameAs ?alignment.}

}

```

}

ORDER BY(?Entity)

Cette requête permet de prendre en compte un maximum de notices, dont certaines qui auraient pu être mal formées à l'origine (par exemple l'absence ou la saisie de qualificatifs non normalisés). De multiples erreurs (plusieurs extérieures à notre corpus) ont ainsi pu être relevées et directement corrigées ou signalées.

Elle permet aussi d'avoir un aperçu plus global de l'ensemble des qualificatifs usités, et de ceux qui pourront potentiellement intégrer le corpus : certains renvoyant à des cultures andines ou asiatiques ont pu être directement évacués, tandis que d'autres ont pu être discutés en interne ou soumis à l'avis du comité scientifique.

Il s'agit d'une requête SPARQL pensée spécifiquement pour le corpus traité : les expressions régulières utilisées en filtres se focalisent sur les qualificatifs et des entrées par termes génériques notamment. Mais elle doit pouvoir être repensée et adaptée selon les besoins d'un autre ensemble : pour cela, il faut évidemment avoir préalablement observé en détail un certain nombre de notices pour se familiariser avec les filtres clés. Notons que l'application Yasgui d'IdRef propose divers exemples de requêtes et permet aussi de retenir un historique.

### **1.1.1.2 Extraction des données de PACTOLS**

Le projet s'est appuyé sur une version du logiciel OpenTheso qui a été revue depuis. Une nouvelle version a été en effet déployée pour PACTOLS le 1er juillet 2020<sup>1</sup>. Pour permettre une meilleure compréhension du travail accompli le vocabulaire et les illustrations employées ci-dessous seront ceux de l'ancienne version. Mais pour mettre en perspective des applications à venir, il convient d'exposer les manipulations équivalentes dans la nouvelle version.

Avec un compte, il est possible d'accéder aux fonctionnalités avancées de PACTOLS dans OpenTheso. L'onglet Édition (*la boîte à outils*) donnait ainsi accès à l'outil d'extraction du thésaurus, dans son intégralité ou par filtre de domaines (*collections*) et de langue.

Dans notre cas, le CSV pro a été exporté (aucun format similaire disponible à l'heure actuelle), en cochant UF pour "Used For" qui permettait d'extraire les variantes de nom. Une fois celui-ci téléchargé, il faut encore le traiter dans OpenRefine pour exclure les termes hors corpus.

La structure du CSV pro était telle qu'elle présentait les différents niveaux de termes génériques sur des colonnes séparées (N1, N2, etc.). Il suffit d'utiliser la fonction "Fill Down" pour pouvoir opérer un facettage et filtrer. OpenRefine offre des outils de traitement, nettoyage et enrichissement d'ensembles de données, ainsi que leur recroisement. Il permet de copier les historiques d'opération sous format JSON sur un fichier et de les reproduire (en les

---

<sup>1</sup> OpenTheso est libre et son code accessible sur Github :

<https://github.com/miledrousset/opentheso>.

Pour une introduction au logiciel, voir la présentation de Miled Rousset lors de la conférence FAIR Heritage: Digital Methods, Scholarly Editing and Tools for Cultural and Natural Heritage (17-18 juin 2020) :

[http://www.lestudium-ias.com/sites/default/files/public/video/Miled\\_Rousset\\_OpenTheso\\_PACTOLS.mp4](http://www.lestudium-ias.com/sites/default/files/public/video/Miled_Rousset_OpenTheso_PACTOLS.mp4)



manipulant). Voici un exemple d'opération de remplissage puis de filtrage utilisée au tout début du traitement du fichier .csv extrait de PACTOLS :

```
[
  "v": {
    {
      "v": "Artistes",
      "op": "core/fill-down",      "l": "Artistes"
      "engineConfig": {      }
      "facets": [],      },
      "mode": "row-based"      {
    },      "v": {
      "columnName": "N1",      "v": "Ecrivains",
      "description": "Fill down      "l": "Ecrivains"
      cells in column N1"      }
    },      },
    {
      {
        "op": "core/row-removal",      "v": {
          "engineConfig": {      "v": "Personnages",
          "facets": [      "l": "Personnages"
          {      }
          "type": "list",      }
          "name": "N1",      ],
          "expression": "value",      "selectBlank": false,
          "columnName": "N1",      "selectError": false
          "invert": false,      }
          "omitBlank": false,      ],
          "omitError": false,      "mode": "row-based"
          "selection": [      ],
          {
            "description": "Remove
          "v": {      rows"
```

```

    "v": "Savants",    }
    "l": "Savants"    ]
}
},
{
  "v": {
    "v": "Antiquité après
l'Antiquité",
    "l": "Antiquité après
l'Antiquité"
  }
},
{
  "v": {
    "v": "Gouvernants",
    "l": "Gouvernants"
  }
},
{
  "v": {
    "v": "Personnages judéo-chrétiens",
    "l": "Personnages judéo-chrétiens"
  }
},
{

```

OpenRefine permet aussi le chargement de fichiers JSON, s'il n'est pas possible d'extraire de fichier au format CSV.

## 1.1.2 Comparaison dans OpenRefine

La comparaison des deux ensembles est réalisable grâce au logiciel OpenRefine. PACTOLS présente habituellement ses entités sans qualificatifs, puisque le terme générique doit lui-même servir de descripteur. Pour éviter la confusion entre deux termes similaires cependant, les entités peuvent se voir accoler un descripteur entre parenthèses, qui intègre directement le nom de l'entité. Ces descripteurs ne sont pas normés (ex : Pallas (géant) / Pyrrha (héroïne)). Il faut comme pour IdRef les écarter.

Un nettoyage poussé demande de vérifier la suppression des espaces, la bonne capitalisation des noms (formules automatisées sous Edit Cells > Common Transforms) ainsi que la suppression d'accents qui pourraient empêcher l'alignement (une regex est ici nécessaire, voire une recette [Jython](#), une autre option est de passer par Dataiku<sup>2</sup> où un outil de traitement des ponctuations et diacritiques est déjà prêt à l'usage).

Pour recroiser des données entre plusieurs jeux chargés dans OpenRefine, il faut utiliser l'outil "Add a column based on this column" sur les colonnes qui vont servir de pivot (le nom et les variantes) et y insérer la fonction suivante :

```
cell.cross('nom de l'autre projet','nom de la colonne à comparer à celle pivot').cells['nom de la colonne dont on veut importer les valeurs'].value[0]
```

Malgré cela, il est encore possible que des entités identiques ne soient pas alignées entre elles. Pour toute entité de l'un ou l'autre vocabulaire, une recherche manuelle peut encore être réalisée pour vérifier qu'il n'y a aucun manquement, et éviter plus tard la production d'un doublon. Certaines entités de nature double ou trouble demandèrent même de revenir au référentiel pour vérifier qu'il n'était simplement pas classé dans une autre branche (PACTOLS) ou avec un qualificatif non pris en compte dans notre recherche (IdRef).

Exemple : Romulus possède une variante dans IdRef avec le qualificatif "mythologie romaine". Cependant, il était absent de la branche héros romains de PACTOLS : il existe mais avait alors pour unique parent "Rois de Rome", un sous-ensemble de "Gouvernants" dans Anthroponymes. Les Horaces, présents dans la branche héros romains de PACTOLS, n'avaient d'abord pas été repérés dans IdRef, où ils étaient qualifiés de "personnages fictifs".

Les différences qui existent entre PACTOLS et IdRef permettent aussi de repérer certaines confusions ou oublis d'autorités homonymes dans l'un ou l'autre répertoire, par des redoublements d'alignement. Ainsi, Hippolyte, entité unique dans IdRef au début du travail, fut aligné une fois avec Hippolyte reine des Amazones et une seconde avec Hippolyte fils de Thésée, les deux existant déjà dans PACTOLS.

## 1.2 Création des notices manquantes

Dans le cas du projet RefDivinités, le travail de complétion s'est fait manuellement, en raison du nombre restreint d'entités manquantes de part et d'autre et suite aux vérifications manuelles déjà réalisées sur un corpus avec risques d'homonymies.

---

<sup>2</sup> Dataiku (<https://www.dataiku.com/>) est un logiciel propriétaire. Cependant, un certain nombre d'outils sont accessibles dans sa version gratuite. Notamment, il permet le suivi de projets et la modification des ensembles de données.

Pour un projet de plus grande envergure et présentant un corpus ne comprenant pas les mêmes risques, il est évidemment possible, et souhaitable, de recourir à des imports de notices en masse.

La création d'entités dans PACTOLS à partir de notices IdRef reste simple, n'utilisant dans un premier temps que le point d'accès autorisé comme entrée.

Dans le sens inverse, une notice d'autorité IdRef devait présenter une ressource bibliographique justifiant l'appellation et une justifiant sa création. L'Abes a souligné que, dans le cadre du projet, la mention de PACTOLS en source était suffisante pour justifier leur production sur la base d'une entité préexistante dans le thésaurus de FRANTIQ.

## 2. Réflexion méthodologique avec les spécialistes

Un comité scientifique a été réuni par Sébastien Dalmon, conservateur chargé de collection en sciences de l'Antiquité à la BIS et chef de projet RefDivinités. Lui-même titulaire d'un doctorat en histoire ancienne (avec un sujet de thèse portant sur la religion grecque), il a établi une liste de chercheurs et chercheuses en sciences de l'Antiquité, plus particulièrement spécialistes des religions. Au total, 20 personnes se sont montrées intéressées et disponibles pour participer à des réunions ou être consultées ponctuellement participer à distance aux débats et répondre à des questions plus spécifiques à leur domaine ([cf. annexe 1](#)).

Les comptes rendus des réunions du comité scientifique font partie de la documentation accessible sur la page CollEx du projet :

<https://www.collexpersee.eu/projet/interoperabilite-de-referentiels-sur-les-divinites-et-heros-du-monde-mediterraneen-antique/>

### 2.1 Familiariser le comité scientifique avec les enjeux

De nombreux spécialistes ont leurs propres ressources et accès. De ce fait, ils n'ont pas forcément conscience, ni de l'existence de certains services, ni des enjeux dans leur gestion et leur amélioration. S'ils connaissent bien le Sudoc, peu connaissent ainsi le site IdRef. Ce travail de sensibilisation, qui recentre et redéfinit certains éléments pouvant toucher la nature même du corpus, est réalisé en parallèle avec le travail d'identification des entités.

Le premier point a été d'établir qu'il ne s'agit pas d'un projet encyclopédique. Notre mission ne tend pas à un recensement exhaustif : l'existence des entités dans IdRef ou PACTOLS est justifiée par leur utilité pour le catalogue.

Il ne doit pas non plus servir de source encyclopédique : les éléments les décrivant servent avant tout à leur bonne identification. Avec ce projet toutefois, nous prenons les devants de besoins potentiels en créant des notices en miroir de celles qui existent dans chacun des vocabulaires et selon l'avis des scientifiques familiers de la documentation. Enfin, cette mission doit servir un public large, de professionnel.les de bibliothèques tout d'abord qui ont la mission de cataloguer et indexer à l'aide de ces autorités, des étudiant.es et chercheur.ses d'autre part qui les utilisent pour explorer les bases de données. Les propositions du comité doivent être adaptées à ces usages et non pas imposer une logique théorique irréconciliable avec certaines conceptions préexistantes.

## 2.2 Évaluation des problématiques avec les spécialistes

### 2.2.1 Fixer le corpus

Un premier ensemble a pu être présenté au comité scientifique en début de mission. Les entités qui sortaient sans ambiguïté du domaine méditerranéen antique avaient déjà été écartées : Amériques, Antilles, Afrique subsaharienne, Asie de l'Est et du Sud-Est, Pacifique, Europe du Nord. Les entités celtiques connues par des sources romaines ont été acceptées. Ont été en revanche exclues les puissances et figures celtiques irlandaises, ainsi que les germaniques et nordiques, attestées uniquement par des sources médiévales. Certaines entités attestées dans des lieux ou par des témoignages correspondant à l'aire culturelle et géographique établie ont pu être intégrées : c'est le cas de puissances de la péninsule arabique présentes à Palmyre. Un travail de désambiguïsation des entités a été mené : des notices pouvant regrouper plusieurs noms et confondre une variété de figures ou puissances ont été distinguées, à l'exemple d'Ashtart (divinité sémitique) issu du vocabulaire de la BnF et dont les variantes de nom présentaient de nombreuses divinités, certaines liées ou assimilées mais d'autres foncièrement différentes (Astarté, Ishtar, Tanit, etc...).

Il a aussi été demandé aux spécialistes de proposer d'autres entités, notamment pour pallier certains manquements ; les corpus hittites ou étrusques ont été développés à cette occasion. Cette demande devait toutefois être accompagnée d'une littérature la concernant, justifiant sa création, bien que la suggestion par le comité soit déjà un premier indice de l'importance de ces entités.

### 2.2.2 Redéfinition des terminologies

Les discussions plus théoriques engagées avec le comité scientifique ont porté sur le vocabulaire structurant l'ensemble du corpus, à savoir la manière dont les entités étaient classées et qualifiées.

#### 2.2.2.1 Terminologies des concepts

Comme évoqué en introduction, PACTOLS distinguait "divinités" et "héros" dans son thésaurus d'Anthroponymes tandis qu'IdRef propose, sur le modèle de la BnF, une séparation entre les qualificatifs de "divinité(s)", classés habituellement sous *Dieux* ou *Déesses*, "mythologie" classés sous le même terme et "personnage légendaire".

Un premier obstacle a été très vite évacué par les spécialistes : sous "Héros grecs" dans PACTOLS existe une catégorie "Créatures de la mythologie grecque". Cette distinction est apparue superflue à leurs yeux, en plus d'apparaître en contradiction avec son générique de "Héros" et le terme a été déplacé pour ne plus servir de sous-catégorie.

Les termes de divinités, héros et mythologie ne leur ont pas non plus semblé refléter l'évolution du vocabulaire dans le monde de la recherche. Un vote a été organisé pour permettre de s'accorder sur un choix de terminologie :

- Puissances suprahumaines : l'expression doit permettre d'embrasser plus largement la variété des entités qui ont fait l'objet de culte et vénération dans le monde méditerranéen antique.

- Figures : le terme a été préféré à personnage pour permettre à son tour de réunir divers types d'entités peuplant l'imaginaire antique, certains objets narratifs d'autres iconographiques, sans distinction et permettant une mise à égalité.

Ces expressions doivent servir dans PACTOLS à catégoriser les entités. Leur obscurité par rapport au langage courant (contrairement aux termes "divinités" et "héros") ne permet cependant pas qu'elles soient elles-mêmes utilisées pour l'indexation d'une part, ni qu'elles puissent figurer comme qualificatifs des mêmes entités dans IdRef.

Une troisième option unique a été envisagée pour ce référentiel : "religion et mythologie" ou simplement "religion". L'avantage de cette expression est qu'elle permet à la fois d'esquiver le terme de divinité, et qu'elle autorise un flou sur la nature des entités, certaines classées dans mythologie mais dont un culte ("héroïque") a pu être attesté, ou de manière générale dont l'inscription dans l'art et la littérature dépasse le cadre purement fictif et narratif associé au mythe. Cette formule, ainsi que d'autres suggestions pour l'évolution de la structure et du vocabulaire dans IdRef, ont été soumises à l'Abes et à l'équipe Rameau de la BnF. Elle sert aussi pour l'instant de terme générique aux deux termes de regroupement dans PACTOLS.

### **2.2.2.2 Terminologies des aires**

La réflexion autour des qualificatifs et des classifications a aussi porté sur les termes définissant l'appartenance d'une entité.

En effet, les termes usités dans les deux référentiels sont de natures variables, marquent des biais anciens de la recherche et des absences importantes. Dans PACTOLS, le terme "oriental" aussi bien pour qualifier un vaste ensemble de divinités ou de héros est ainsi extrêmement problématique.

Aussi, une réflexion sur la terminologie et son arborescence a été menée durant l'été 2019, pour permettre de classer et qualifier avec plus de pertinence les entités du corpus. Le choix a été fait de formules qui permettent d'articuler des réalités géographiques et chronologiques, embrassant des concepts contemporains sans qu'il y ait toujours de correspondance possible avec des religions aux limites clairement définies.

Ce tableau ([cf. annexe 2](#)) est à apprécier en fonction des aires qui nous ont concerné. Son implémentation partielle dans PACTOLS et son appropriation pour la qualification de notices IdRef doivent être prises comme des incitations à réfléchir à un classement plus large, mais qui doit rester souple et à l'écoute d'avis divergents.

## **2.3 L'établissement des sources**

La *source* est un élément essentiel dans IdRef pour justifier non seulement la création d'une entité dans le référentiel, mais aussi l'ensemble des informations qui la décrivent.

Le guide du catalogueur recommande tout d'abord une référence générale, notamment pour justifier le choix du point d'accès autorisé, qui doit être en français ou d'usage courant, pour les "divinités et héros".

Aussi, nous avons réuni un ensemble de ressources touchant à différents niveaux de spécialisation. Nous avons aussi demandé au comité scientifique que leurs suggestions d'autorités à ajouter au corpus soient accompagnées d'une ressource bibliographique dont

elle est sujet. Un même travail de justification leur a été demandé pour leurs suggestions de corrections sur les notices existantes.

Certaines entités préexistantes possèdent une ou plusieurs sources bibliographiques, celles-ci ont été conservées dans la mesure de leur pertinence. Il peut s'agir de références encyclopédiques (*Grand Larousse Universel*, *Encyclopedia Universalis*) ainsi que de ressources spécifiques.

Pour le reste, une hiérarchie a été établie comme suit :

- Le dictionnaire *Dieux, Déesses, Démons* (DDD) sous la direction de Patrick Jean-Baptiste (Paris : Seuil, 2016) a été établi comme première ressource pour son caractère englobant, généraliste et sa publication récente.
- En deuxième lieu, une ressource encyclopédique spécialisée est utilisée, qui doit pouvoir servir de ressource encyclopédique si l'autorité n'a pas de source préexistante et qu'elle ne figure pas dans le DDD : le *Dictionnaire de la mythologie grecque et romaine* de Pierre Grimal (Paris : P.U.F., 1951) pour les religions et mythologies grecques et romaines, le *Dictionnaire de la mythologie et de la religion celtiques* de Philippe Jouët (Fouesnant : Y. Embanner, 2012) pour les mythologies et religions celtiques ;
- Des ouvrages spécialisés divers peuvent figurer, notamment en l'absence de ressource encyclopédique spécialisée francophone (*Dieux et déesses de l'univers phénicien et punique* d'Edward Lipinski (Leuven : Peeters, 1995) pour la Phénicie par exemple).
- Enfin, une dernière ressource encyclopédique, cette fois-ci d'usage international, peut être utilisée : *Brill's New Pauly* (Leiden : Brill, 2002-2010), *Reallexikon der Assyriologie* (Berlin : W. de Gruyter, 1928-...).

La dernière référence bibliographique figurant dans la notice d'autorité est habituellement celle justifiant sa création et son utilité. Il peut s'agir d'un ouvrage voire d'un article (qui dans ce cas n'est pas indexé dans le SUDOC). Il n'y en a pas forcément pour les notices préexistantes, ce qui ne pose pas problème pour les autorités riches en indexation. Pour celles que l'on crée, le lien vers PACTOLS est suffisant (voir 1.2).

Ces ressources fournissent les éléments de description de la notice, mais à plusieurs reprises ces indications ont pu être revues par des spécialistes du comité scientifique.

## 2.4 Maintenir le dialogue

Un travail non négligeable a été investi dans l'élaboration et la tenue des réunions du comité scientifique. Toutefois, il est impossible de faire entendre toutes les voix ni de toucher à tous les sujets durant ces quelques sessions.

Entre les 2<sup>e</sup> et 3<sup>e</sup> réunions du comité, un questionnaire a été envoyé, à remplir par chaque membre du comité. Celui-ci a permis de prendre des décisions par vote sur la terminologie, ainsi que de récolter les suggestions de classification et de nouvelles entités, mentionnées plus haut.

Par la suite, certains membres du comité ont été individuellement interrogés sur des points précis, sur une partie spécialisée du corpus, voire sur certaines autorités particulières. La difficulté à mobiliser ces chercheurs et chercheuses sur leur temps libre est à souligner, et il faut garder à l'esprit que leur collaboration gratuite et volontaire ne peut avoir requis leur attention et expertise que partiellement. Ce projet comme tant d'autres est non seulement partiel mais encore perfectible.

### 3. Traitement des entités dans IdRef

Il est possible de créer et manipuler des entités dans IdRef de plusieurs manières dès lors que l'on possède des identifiants fournis par l'Abes. La première est via la plateforme IdRef elle-même. La seconde est grâce au logiciel WinIBW. La troisième enfin, plus spécifique aux demandes de création ou de changement en masse, se fait par requête auprès de l'Abes.

Outre les recommandations de l'Abes et après nous être entretenus avec elle, le réseau des correspondants autorités du Sudoc a été contacté pour obtenir certaines suggestions et éclaircissements sur la structure type et complète à envisager pour les notices de notre corpus.

#### 3.1 Traitement pratique des notices

Il est nécessaire de créer un script dans WinIBW<sup>3</sup> selon un schéma déterminé. Cela permet d'homogénéiser les structures, de garder à l'esprit la mention d'informations et plus simplement d'automatiser une partie du processus de création. Le script présenté ci-dessous est la matrice à partir de laquelle ont pu être créées une multitude d'entités pour le projet RefDivinités.

```
Sub CreerTpRefDiv ()
```

```
'Ce script permet de créer une notice d'autorité personne physique Tp5 pour le projet RefDiv application.activeWindow.codedData = false application.activeWindow.command "create", false
```

```
Application.ActiveWindow.Title.InsertText"008 $aTp5"&vblf&_
```

```
"033 ##$ahttp://catalogue.bnf.fr/ark:/12148/Identifiant BnF commençant par "cb"$2BNF$dYYYYMMDD" &vblf&_
```

```
"035 ##$aFRBNF+9 signes$zFRBNF+8 signes" &vblf&_
```

```
"035 ##$ahttps://viaf.org/viaf/Identifiant VIAF/$2VIAF" &vblf&_
```

```
"035 ##$ahttps://ark.frantiq.fr/ark:/26678/n°ark$2PACTOLLS$CFRANTIQ$dYYYYMMDD" &vblf&_
```

```
"035 ##$aIdentifiant Wikidata (Q+série de chiffres)$2WIKIDATA$dYYYYMMDD" &vblf&_
```

```
"106 ##$a2$b1$c0" &vblf&_
```

```
"120 ##$ax" &vblf&_
```

```
"200 #0$8frefre$90y$aNom$cQualificatif de fonction"&vblf&_
```

```
"340 ##$aNote biographique"&vblf&_
```

---

<sup>3</sup> Pour plus d'informations sur la création de scripts, voir la documentation de l'Abes : [http://documentation.abes.fr/sudoc/manuels/logiciel\\_winibw/scripts/index.html#ChoisirScripts](http://documentation.abes.fr/sudoc/manuels/logiciel_winibw/scripts/index.html#ChoisirScripts)



"400 #0\$9#y\$aVariante de nom\$cQualificatif de fonction"&vblf&\_

"500 #0\$3"&vblf&\_

"550 ##\$5g\$3"&vblf&\_

"686 ##\$aRefDivinités\$cAppel à projets 2019\$2Collex-Persée"&vblf&\_

"810 ##\$aRessource encyclopédique\$bInformation obtenue"&vblf&\_

"810 ##\$ahttps://pactols.frantiq.fr/opentheso/, YYYY-MM-DD"&vblf&\_

"822 12\$aNom\$2LCSH\$uhttp://id.loc.gov/authorities/subjects/ID Subject\$dYYYY-MM-DD"&vblf

Notons qu'une sous-zone mal ou non remplie provoquera une erreur qui empêchera la sauvegarde de la notice tant que l'erreur n'aura pas été rectifiée, permettant d'éviter des saisies erronées.

Les zones 033 et 035 d'alignement avec des vocabulaires extérieurs sont sauvegardées et inchangeables dès validation de la création ou de modification de la notice, ou bien de la translittération automatique. De nouvelles zones 035 peuvent être ajoutées mais il faut demander au correspondant Abes de notre institution pour toute autre modification. Il est à signaler que ces précautions ne se limitent pas à la saisie manuelle de ces zones mais doivent aussi être prises lors de la soumission d'un tableau à l'Abes pour enrichissement automatique : la structure des adresses et zones doit être anticipée.

Les zones 106 et 120 sont spécifiques à notre type d'entités, qui ne peuvent pas être autrices mais uniquement sujets, et qui, par leur nature, dépassent la possible classification de sexe. De manière similaire, n'étant pas *nées* dans un pays et n'ayant aucune langue, ces éléments ne sont pas référencés et ont été supprimés des notices qui les mentionnaient.

Celles créées durant notre projet ont varié à partir de cette base, présentant moins d'alignements (033 et 035), des compléments d'information sur la langue (\$8) ou la translittération (\$9) en 4XX, une possible entrée en 7XX, une alerte en 820 à destination des catalogueurs pour éviter des confusions et une 822 permettant l'alignement avec la LoC (voir ci-dessous).

### 3.2 Normes fixées

La création et l'enrichissement de notices dans IdRef répond à un ensemble de normes fixées par la BnF et l'Abes. Pour répondre à ces normes, le guide Abes et le guide du catalogueur de la BnF sont à notre disposition<sup>4</sup>. En cas de problématique insoluble, le guichet de l'Abes doit permettre d'apporter les précisions demandées.

---

<sup>4</sup> Guide Abes sur la translittération :  
[http://documentation.abes.fr/sudoc/doc/Formateur\\_relais/Translitteration\\_multiecritures\\_dans\\_le\\_Sudoc.pdf](http://documentation.abes.fr/sudoc/doc/Formateur_relais/Translitteration_multiecritures_dans_le_Sudoc.pdf) ;

Guide du catalogueur de la BnF sur la translittération :  
[http://guideducatalogueur.bnf.fr/abn/GPC.nsf/gpc\\_page?openform&type\\_page=fiche&unid=671104F78A605324C12574DC002F04ED#rem2](http://guideducatalogueur.bnf.fr/abn/GPC.nsf/gpc_page?openform&type_page=fiche&unid=671104F78A605324C12574DC002F04ED#rem2)

### 3.2.1 Normes de translittération

Notre corpus est composé pour près de la moitié d'autorités de la mythologie ou de la religion grecque, aussi, la translittération à partir de l'alphabet grec est un enjeu important. La translittération est obligatoire dès lors qu'un point d'accès est dans un alphabet non latin.

L'Abes utilise la norme ISO 843 pour la translittération du grec ancien. Dans l'application WinIBW, cette norme est celle (supposément) utilisée lorsqu'on fait appel aux outils de translittération automatique.

Dans certains cas, la norme ne correspond pas à la tradition érasmiennne. Une correction est donc apportée en variante à la translittération donnée, identifiée par la sous-zone : \$9#d.

Lettre	Translittération ISO 843	Renvois à faire
φ	f	ph
β	v	b
η	ī	ē
κ	k	C (en début de mot)
χ	ch	Kh (en début de mot)

À noter que l'ensemble des renvois à faire sont réunis dans une même variante de forme.

Les ressources utilisées ainsi que les usages fournissent un certain nombre de variantes. Certaines ne suivent pas la norme ISO 843, c'est le cas de la *New Pauly* qui se base sur les normes de translittération américaines de l'ALA-LC. Dans le cas où ces formes se distinguent de la translittération ISO 843 et du renvoi potentiel qui en découle, elles sont présentées en zone 4XX avec le code de sous-zone \$9#f.

Il est alors nécessaire de préciser la nature de la translittération en zone 340.

Ex : 340 ##\$aPaiēlōn : translittération ALA-LC.

D'autres formes hybrides comme le LIMC ou l'*Année philologique* ne peuvent être considérées comme de réelles translittérations. Leur apparition en variante ne doit pas s'accompagner d'une sous-zone \$9, ou bien sous la forme \$9#y.

### 3.2.2 Prendre en compte les évolutions du guide méthodologique et leur impact potentiel

Les mises à jour signalées par l'Abes, certaines à la suite de discussions avec eux, ont permis d'apporter un nouvel éclairage sur certaines réflexions.

Depuis les recommandations du 13 septembre 2019 instaurées par le [Guide de l'Abes, la zone A035](#) présente de nouveaux exemples de structuration pour les alignements (Wikidata, PACTOLS, etc.) et invite à en proposer d'autres suivant les besoins et projets.

À la suite des recommandations du 24 octobre 2020, la question des formes en écriture d'origine ou translittérée en zone A2XX et A7XX soulève deux points en ce qui concerne les entités mythiques et divines.

Selon le guide du catalogueur de la BnF, la forme du point d'accès autorisé, la zone 2XX, doit être une forme courante, française si elle existe. Pour les personnes réelles, il peut y avoir une 2XX dans son écriture d'origine. Le guide précise que des formes « parallèles et/ou exclues en caractères non latins » peuvent être admises, comme l'[exemple de la divinité hindoue Sītā](#) le souligne. Il faut donc faire la part entre les formes à inscrire en 7XX et celles en 4XX. La forme originelle en alphabet non latin peut figurer en 7XX, accessible dans des dictionnaires comme Grimal pour la langue grecque, avec sa translittération obligatoire. En 400, nous retrouverions donc diverses variantes, formes non normées, voire d'autres formes en alphabet non latin et leurs translittérations.

Le second point est celui des sous-zones \$8 (langue de catalogage et langue de la racine du point d'accès) et \$9 (valeur de la forme et système de translittération).

Dans le guide du catalogueur, on peut observer des notices où seules les positions 0-2 de la \$8 sont remplies, indiquant la langue de catalogage (le français, sauf dans le cas où les qualificatifs seraient importés d'un catalogue non francophone et entrés dans une autre langue). Dans IdRef, la présence de 6 caractères est nécessaire, interdisant qu'on ne précise que la langue de catalogage. Les positions 3-5 précisent la « langue racine du point d'accès ». Mais selon les nouvelles directives RDA-FR, il s'agit de la « langue d'expression » de la personne, ce qui n'est évidemment pas pensable pour les entités de notre corpus. Il faut donc se satisfaire de la règle qui veut que cette position désigne la langue de la forme du nom.

La sous-zone \$9 n'est jamais précisée dans le guide du catalogueur. On peut supposer, comme la nature de la 200 doit être une forme d'usage courant en France, qu'il faille que position 0=1, mais ce n'est possible que si l'entité ne présente qu'un point d'accès autorisé, ou plus de deux (ce qui est techniquement le cas si un couple d'une forme en caractère non latin et sa translittération est répertorié en zone 700). Toutes les 400 présentent une position 0=#. En position 1, le système de translittération doit être précisé. La valeur "a" est à utiliser en fonction du type de translittération effectuée. Autrement, la valeur "y" sera principalement utilisée, qui précise qu'il ne s'agit pas d'une translittération.

Avec les recommandations du 18 novembre 2019 sur les noms de famille (A220/A420), la sous-zone \$c est désormais réservée à la précision du « type de famille ».

Celle-ci peut être décrite de différentes manières : famille / clan / dynastie / frères / sœurs.

Une sous-zone \$d peut préciser la localisation de cette « famille » (pays / ville etc.).

Cela pose la question de la gestion des familles fictives. Doit-on en faire des personnes (Tp) ou des noms communs (Td) ? Peut-on malgré tout utiliser la \$c pour préciser la nature fictive ou mythologique de la famille (en un sens, on pourrait souligner qu'il s'agit d'un type de famille) si les entités sont maintenues en Ta ? La \$c en zone 200 correspond à tout « élément additionnel autre que les dates » ou « qualificatif de fonction » et c'est ainsi qu'il était vraisemblablement utilisé en 220 jusqu'à présent. Il a été choisi pour les Atrides et les Labdacides de proposer une sous-zone : \$cfamille ; mythologie grecque.

De manière plus large, la notion de famille et des entités plurielles peut sembler confuse, on retrouve en Td et Tp des groupes de sœurs, frères, couples (Danaïdes, Dioscures, Philémon et Baucis, Pyrame et Thisbé). Cette interrogation a été soumise à l'Abes et RAMEAU et s'inscrit dans la réflexion de la Transition bibliographique.

Enfin, depuis le 7 janvier 2020, la zone A822 est structurée pour accueillir un alignement avec un thésaurus extérieur, Laval RVM ou bien de la Bibliothèque du Congrès (LoC). Cette zone permet en outre de préciser l'absence ou l'invalidité d'un alignement. Cela peut amener à déplacer en 810 ou 815 certaines informations qui relevaient de ces vocabulaires.

### **3.2.3 Suivi des modifications**

Il est possible de suivre la création des nouvelles entités en comparant la liste originelle des notices et celle des nouvelles notices créées. Il en va de même pour les informations des différents champs d'importance préalablement listés. La structure des notices doit être identique pour toutes, sauf exceptions discutées. Ces informations sont dans la plupart des cas absentes des notices préexistantes. Dans le cas où elles existaient, le suivi de leur modification doit malgré tout être connu en interne.

Concernant les champs libres, particulièrement les zones biographiques et de sources : ces zones doivent être maintenues tant que leur suppression ou modification n'est pas justifiée.

## **3.3 Signalement des problématiques à la BnF/Rameau et à l'Abes**

Si les entités existent dans le catalogue de la BnF, il est nécessaire de vérifier leur correspondance exacte avec la notice qui leur serait potentiellement associée. Dans le cas de notices importées qui auraient été modifiées (dédoublonnage, réévaluations), ces modifications doivent être signalées à la BnF pour permettre une révision de leurs propres notices. On retrouve dans notre corpus un certain nombre d'erreurs d'indexation (doublons, mauvaises catégorisations, etc.) au sein de ce catalogue.

Une documentation spécifique a été rédigée à l'intention de RAMEAU, de l'Abes et de la BnF pour signaler les problématiques rencontrées. Les autorités RAMEAU dans IdRef ne sont pas modifiables à notre niveau : cela concerne non seulement les termes de classification (Mythologie, Dieux, etc.), mais aussi un ensemble que nous avons désigné comme « entités plurielles ». Les débats avec le comité scientifique ont permis d'avancer des propositions de changement de dénomination et de classification de ces premiers termes, mais cela demande évidemment une réflexion plus approfondie des agences, ainsi que la prise en compte plus globale des enjeux de la transition bibliographique.

Concernant les « entités plurielles », leur nature confuse a été soulignée : certaines sont des Td (noms communs) généralement issues et sous l'autorité de RAMEAU comme les Muses, les Centaures, d'autres sont des Tp, noms de personnes gérés par l'Abes comme les Nymphes ou les Titans. Les « noms communs » n'ont ainsi pas pu être modifiés. Le débat peut même être prolongé sur ce qu'ils englobent : les cyclopes ou centaures désignent-ils exclusivement ceux de la mythologie grecque et devraient-ils être alignés avec ceux de PACTOLS ?

Ces questionnements ont accompagné le projet, et ont été portés à l'attention de la BnF/RAMEAU et de l'Abes, pistes ouvertes par RefDivinités mais qui mènent à des enjeux plus larges (voir 5.2).

## 4. Alignement des données

L'alignement d'IdRef et PACTOLS est au cœur du projet RefDivinités, mais ne fonctionne pas en vase clos. Pour une meilleure intégration de ces autorités au web de données, plusieurs autres pistes d'alignements ont été envisagées.

### 4.1 Quels alignements pour quels enjeux ?

#### 4.1.1 BnF

Les autorités de la BnF possèdent un identifiant ark unique. La plupart des entités déjà existantes de notre corpus qui ont un équivalent dans la base de la BnF ont un alignement déjà partiellement établi (notamment avec un identifiant FRBNF distinct de l'ark). Certaines entités au contraire n'ont pas été directement importées de la BnF voire encore créées dans IdRef.

Il a été établi que ces dernières aussi auraient bénéficié à être créées à partir du moment où elles existaient déjà dans les autorités de la BnF, à la manière des entités déjà existantes dans PACTOLS. Le projet RefDivinités a donc importé l'ensemble des entités "divinité / mythologie" touchant au corpus dans IdRef et complété les alignements absents ou partiels.

#### 4.1.2 VIAF

VIAF est une base de données gérée par l'OCLC. Son but est de réunir l'ensemble des notices internationales correspondant à une même autorité. Il permet de faire la liaison entre les bibliothèques nationales, mais aussi d'autres projets comme Wikidata. Dans le cas où une autorité possède une notice dans IdRef et dans la BnF, celles-ci sont habituellement liées, car le lien était présent et l'alignement aisé.

Cependant, au sein de notre corpus et à plus grande échelle pour les entités fictives, les divergences sont une quasi constante entre ces identifiants de la BnF et IdRef ainsi qu'avec d'autres notices internationales. Pour une même puissance comme Artémis, on retrouvera par exemple une notice de la Bibliothèque nationale d'Allemagne et une d'IdRef chacune isolées, puis une de la Bibliothèque nationale d'Israël liée à celle de la Bibliothèque du Congrès américaine. Il faut donc envisager un travail de signalement auprès d'OCLC et de fusion de ces facettes doublées.

Dans le cadre de notre travail, il a été décidé de retenir l'identifiant VIAF de la facette IdRef lorsqu'il existait, ou de retenir celui existant le plus ancien (LoC ou DNB la plupart du temps) s'il en existait plusieurs internationaux distincts à sa création. En effet, dans l'éventualité d'un dédoublement de ces notices, le VIAF les fusionnerait sur la notice la plus ancienne dont l'identifiant serait conservé.

### 4.1.3 Wikidata

Wikidata offre une autre perspective sur cet alignement. En tant que plateforme libre, son enrichissement passe par des actions individuelles et de *bots*. La base permet aussi les requêtes SPARQL et de nombreux outils son enrichissement en masse<sup>5</sup>. Ses informations à valeur encyclopédique sont structurées, jusqu'à la source de chaque affirmation.

Elle peut servir à son tour de plateforme où convergent les identifiants d'une multitude d'institutions. Durant l'avancée du projet, les entités Wikidata existantes ont été alignées de manière exhaustive avec les autorités IdRef et PACTOLS, et les identifiants IdRef et PACTOLS ont à leur tour été versés dans Wikidata.

### 4.1.4 Library of Congress

La problématique posée par la Bibliothèque du Congrès est double.

D'une part, il s'agit d'un lien dont l'existence est ancienne dans les notices BnF et celles d'IdRef créées à partir de là. La zone A822 est exploitée pour la LoC ainsi que pour le thésaurus de l'université de Laval (qui n'est pas numérisé en accès libre, voir 3.2.2).

D'autre part cependant, il est à noter que la LoC migre certaines entités "deity" et "mythology" de la catégorie *Subjects Headings*, celle normée dans la mise à jour méthodologique de l'Abes, vers la catégorie *Name Authority*. L'évolution de la nature de ces entités risque d'impacter leur structure à l'avenir, et de se retrouver dans celles de notre corpus.

### 4.1.5 Thésaurus du Getty, AAT et IA

L'*Art & Architecture Thesaurus*, support des projets du Getty, réunit une typologie d'objets, d'activités, matériaux, techniques, et d'agents ; le *Getty Iconography Authority* présente un vocabulaire pour la description iconographique d'œuvres et objets patrimoniaux, des motifs aux sujets narratifs et personnages représentés<sup>6</sup>. L'alignement avec ces thésaurus iconographiques est particulièrement intéressant pour PACTOLS dont le vocabulaire spécialisé s'intéresse aussi aux formes et motifs. Notre corpus compte plusieurs entités dont on connaît des représentations et qui apparaissent ainsi dans l'*Iconography Authority*. L'ambition d'utiliser les notices d'autorité IdRef au-delà d'une indexation dans le SUDOC mais aussi pour l'indexation iconographique et de documents dans Calames par exemple (de la même manière que le vocabulaire de la BnF concerne aussi des images) justifierait aussi cet alignement. Aucun alignement n'a été pour l'instant réalisé avec ces thésaurus.

---

<sup>5</sup> Notons qu'un outil Mix'n'Match existe pour aider à l'alignement semi-automatisé de l'ensemble des entités PACTOLS dans Wikidata : <https://mix-n-match.toolforge.org/#/catalog/2700>

<sup>6</sup> L'ensemble des vocabulaires du Getty sont à explorer à cette adresse : <http://www.getty.edu/research/tools/vocabularies/index.html> et sont accessibles à la réutilisation au sein du LoD au format XML et via des APIs.

Les alignements privilégiés dans IdRef ont donc été : la BnF (zones UNIMARC 033 et 035), le VIAF, Wikidata et PACTOLS (zones UNIMARC 035) ainsi que la Library of Congress (zone 822).

## 4.2. Traiter l'alignement des données

Quelle que soit l'approche dans le travail d'alignement des données, celui-ci doit être supervisé par un tableau maître et des exports qui permettent de vérifier son avancée et sa bonne complétion ainsi que de potentielles erreurs d'indexation (coquilles, doublons) repérables de manière plus ou moins efficace par des logiciels d'édition de texte comme Atom.

Un outil intéressant à signaler dans cette perspective est la "[réconciliation](#)" permise par le logiciel OpenRefine. Celui-ci permet d'associer la valeur de chaque cellule de la colonne sélectionnée à une autorité dans une base de données extérieure accessible par une API Rest pour extraire des données structurées. S'il est possible de réconcilier avec VIAF ou la BnF, pour en tirer des identifiants, son utilisation est particulièrement avancée avec Wikidata.

### 4.2.1 Traitement manuel dans IdRef

L'ajout d'alignements dans IdRef a déjà été partiellement abordé plus haut. Dans le cadre de notre projet, cet alignement a été opéré de manière manuelle dans WinIBW, appuyé cependant par le modèle de script pour garder à l'esprit l'homogénéité des notices et veiller à ne pas oublier un alignement potentiel.

Cette stratégie avait plusieurs objectifs :

- S'assurer de la bonne correspondance d'une entité PACTOLS pour chaque entité IdRef alignée.
- Dans le cas d'entités créées en miroir de celles existant dans PACTOLS ou suivant la suggestion du comité scientifique, cela permettait de créer la notice dans son intégralité pour qu'elle soit immédiatement optimisée.
- Dans le cas de notices d'autorité qui existaient déjà, cela permettait de réviser l'ensemble des métadonnées qui les constituaient et de les corriger dans le même temps.

### 4.2.2 Traitement automatisé dans IdRef

Aligner les données de manière automatisée dans IdRef permet de se concentrer sur la création et la complétion d'un tableau, plutôt que sur la modification des notices individuelles. Pour arriver à le remplir cependant, plusieurs étapes sont nécessaires ainsi qu'un travail redoublé de vérification.

Un premier moyen d'obtenir les identifiants est d'effectuer des requêtes, notamment auprès de data.bnf.fr et de Wikidata. Data.bnf peut proposer, outre ses propres notices, des alignements avec VIAF, la Bibliothèque du Congrès et d'autres institutions internationales. Wikidata aussi permet, outre ses propres items, d'accéder à une variété d'identifiants dont VIAF, la LoC, la BnF etc. Mais ces requêtes, en particulier dans le cadre de notre projet comme souligné plus tôt, sont extrêmement parcellaires et potentiellement parsemées d'erreurs. L'identifiant VIAF d'IdRef peut être isolé et plusieurs doublons peuvent exister. Celui indexé dans Wikidata pourrait être un différent, et celui de la LoC ne pas être à jour.

Il faut donc veiller à bien aligner les entités PACTOLS et IdRef entre elles, et correctement les associer aux identifiants extérieurs supplémentaires. Une fois ce tableau rempli de manière exhaustive et formaté pour être directement compilable par l'Abes, il est possible de le leur envoyer par l'intermédiaire du/de la correspondant·e autorité de l'établissement pour un import en masse de ces données dans les notices IdRef.

### 4.2.3 Réaliser un alignement dans PACTOLS



Capture d'écran des options d'alignement dans OpenTheso1

Un traitement manuel ou “semi-automatisé” est possible dans le thésaurus de Frantiq. En effet, une fois connecté·e, l'utilisateur·ice peut accéder aux fonctionnalités de traitement des alignements.

Les seules informations structurées sont la nature de l'alignement (<skos:exactMatch>) et son URL. Dans la version 1, il était cependant rendu lisible par l'indication du nom de la source (Wikidata, BnF) ainsi que par un intitulé au concept.

Corpus lié	
Alignement	<hr/> <code>exactMatch:</code> <a href="https://www.idref.fr/241637856">https://www.idref.fr/241637856</a> <code>exactMatch:</code> <a href="http://www.wikidata.org/entity/Q6720682">http://www.wikidata.org/entity/Q6720682</a> <hr/>

Capture d'écran des options d'alignement dans OpenTheso2

L'alignement automatique permet quant à lui d'utiliser une API Rest pour trouver des associations potentielles qu'il suffit ensuite de sélectionner. Il est aussi possible, en sélectionnant le terme générique, de demander l'alignement de l'ensemble des termes spécifiques, chacun demandant confirmation. Il est par ailleurs possible à partir de certaines API d'importer des informations supplémentaires, plus particulièrement avec Wikidata (définition, traductions, images) ou Geonames (coordonnées).

### 4.2.4 Traitement automatisé dans PACTOLS

L'unique scénario actuel qui nous est connu pour un traitement pleinement automatisé de l'alignement dans PACTOLS nécessite qu'il soit envisagé en amont pour un corpus de données non existant dans le thésaurus. Il n'y a pas à cette heure d'outil d'enrichissement en masse de notices existantes.

Sous un terme générique, il est possible de demander : “Importer une liste de termes spécifiques au format tabulé”. Chaque ligne de ce fichier .csv représente une entité et chaque colonne une métadonnée de celle-ci. On peut alors avoir une colonne “skos:exactMatch”

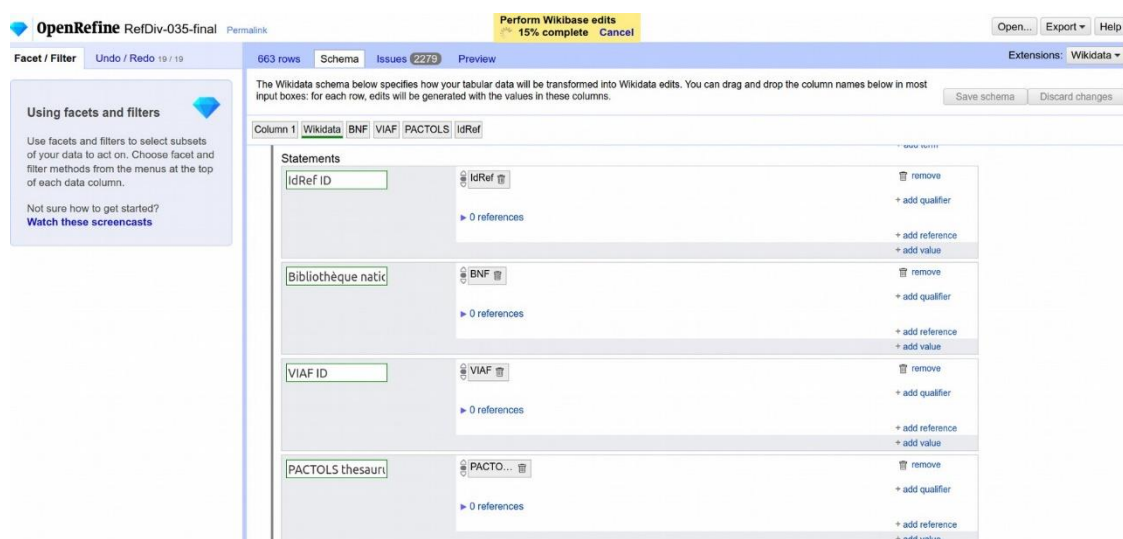


contenant les alignements souhaités qui seront importés en même temps que sont créées ces entités.

Cette stratégie n'a pas été retenue pour notre projet pour plusieurs raisons déjà évoquées, notamment l'existence d'une grande majorité des entités de notre corpus, l'ajout progressif de nouveaux termes, le traitement privilégié des autorités IdRef et la création concomitante des entités dans PACTOLS pour un meilleur suivi.

#### 4.2.5 Traitement automatisé vers Wikidata

Le logiciel OpenRefine offre une panoplie d'outils fondamentaux pour le traitement en masse de données sur Wikidata. Aussi a-t-il été possible de charger directement depuis un fichier d'export IdRef les identifiants d'alignement déjà réunis dans IdRef. Du fait du traitement en masse des données et de la nécessité d'avoir réuni, nettoyé et vérifié les identifiants dans IdRef et PACTOLS, ce travail doit être envisagé en fin de mission.



Capture d'écran du schéma d'identifiants du corpus lors de son chargement dans Wikidata.

L'API particulière permettant cela propose la création d'un schéma wikidata : suite à la réconciliation parfaite de chaque entité avec son identifiant wikidata, le schéma permet de faire de chaque valeur des colonnes normalisées un *item* (ou *objet* dans le langage de graphe rdf) d'une *propriété* (ou *prédicat*) se référant à l'élément wikidata associé. Par exemple, les identifiants PACTOLS déjà alignés avec un identifiant Wikidata dans les thésaurus IdRef et PACTOLS sont traités pour former une *déclaration* (ou *triplet*) dans Wikidata. Une fois toutes les vérifications réalisées, l'ensemble de ces données structurées peuvent être directement chargées dans Wikidata, ou via l'outil Quick Statements.

Il faut noter qu'un compte Wikidata est nécessaire pour effectuer une telle tâche. Ces déclarations seront par la suite modifiables par n'importe qui. Dans la capture d'écran, les *issues* signalées correspondent à l'absence de références qui doit habituellement accompagner l'ajout d'informations à une notice Wikidata. Mais la base est plus souple que l'encyclopédie Wikipédia, d'autant plus que le lien est lui-même référence.

### 4.3 Phases pratiques de traitement durant le projet RefDivinités

Le projet RefDivinités a été ponctué de rencontres avec les spécialistes, de l'émergence de questionnements et fonctionnalités nouvelles, ainsi que de la préparation à des enjeux ultérieurs. D'autre part, la tâche d'enrichissement et d'alignement des données, ainsi que de dialogue et de veille, était réalisée par la même personne, appuyée par les porteurs du projet. Aussi, le travail a été mené par phases, tâtonnements et allers-retours.

La première étape fut celle d'extraction et comparaison des données (voir 1.). Puis, il a fallu présenter ce corpus et réfléchir à son remodelage sémantique, son enrichissement et son extension auprès du comité (voir 2.).

Il y avait en avril 2019 un total de 454 entités identifiées dans IdRef et 463 entités dans PACTOLS.

L'enrichissement d'IdRef a concentré le travail de recherche, l'élaboration des notices, des variantes, alignements et définitions, dans l'optique de leur basculement vers PACTOLS. Ont d'abord été créées dans IdRef les notices qui existaient dans PACTOLS. Puis ont été créées dans PACTOLS les notices IdRef mises à jour manuellement par l'ajout des alignements, éléments descriptifs et ressources. Hors des ouvrages généralistes consultés, il a fallu réaliser un travail important de recherche et consultation de publications spécialisées pour pouvoir enrichir d'un certain nombre d'éléments et justifier la pertinence de l'indexation de multiples entités.

En chiffres : en juin 2020, 664 entités forment l'ensemble de notre corpus dans IdRef (à celles-ci s'ajoutent dans PACTOLS les entités dont les équivalents IdRef sont des Td qui n'ont pas pu être traitées, 7 sont présentes en supplément dans le tableau final). Ces ajouts sont le fait des enrichissements d'un vocabulaire vers l'autre, des suggestions du comité scientifique et d'apports de la BnF.

Il nous a paru pertinent du fait de la gestion du projet de créer au fur et à mesure les entités aussi bien dans IdRef que dans PACTOLS pour permettre un alignement immédiat dans IdRef. Suivant la manière dont une future mission serait organisée, il peut être envisagé de ne pas créer les entités PACTOLS pour compléter les alignements mais d'attendre d'avoir terminé le travail d'enrichissement d'IdRef pour pouvoir importer dans PACTOLS un set de données des nouvelles entités. On enverra consécutivement à l'Abes un tableau d'import en masse d'alignements à partir des notices PACTOLS déjà existantes (préalablement alignées de manière semi-automatique avec IdRef) et nouvellement créées (qui auront été créées avec l'alignement).

Le travail de révision et de restructuration du thésaurus PACTOLS a animé la dernière phase de la mission, prenant en considération les propositions du comité scientifique ainsi que la réflexion des groupes de travail du BBT de DARIAH (voir 5.2.2).

## 5. Valorisations et perspectives autour du projet

### 5.1 La transition vers PACTOLS 2 : des problématiques spécifiques

Le chantier de PACTOLS 2 n'est pas encore achevé. Le projet de réorganisation, lancé en 2017 et soutenu par le consortium MASA (Mémoires des Archéologues et des Sites Archéologiques), doit aboutir d'ici 2022 à une version totalement renouvelée du thésaurus. L'objectif est d'en faire un outil du web des données, disponible à tous pour faciliter l'interopérabilité et la qualification des données en ligne de l'archéologie. Il s'appuie pour cela sur les dernières fonctionnalités du gestionnaire OpenTheso et sur l'ajout d'identifiants pérennes (Ark et Handle) à chaque concept. Du point de vue lexical, des vocabulaires de spécialité sont en cours d'ajout pour équilibrer certains domaines encore peu représentés, ainsi que des alignements aux notices de Wikidata et à d'autres référentiels généraux (GéoNames pour les Lieux) ou du domaine (Art and Architecture Thesaurus). La structure originelle éponyme en micro-thésaurus est quant à elle bouleversée par le modèle du Backbone Thesaurus (BBT) de DARIAH, beaucoup plus conceptuel<sup>7</sup>. Le projet RefDivinités lui-même est à considérer comme un coup d'essai dans le traitement d'un corpus spécifique hors du cadre des groupes de réflexion de PACTOLS, qui doit ouvrir des pistes et lui-même rester ouvert à de nouvelles évolutions.

#### 5.1.1 Propositions pour DARIAH

En se basant sur le modèle du BBT, plusieurs structurations ont été envisagées et transmises à PACTOLS pour faire évoluer les figures et puissances suprahumaines hors du micro-thésaurus Anthroponymes. Evelyne Sinigaglia est intervenue au nom de Frantiq auprès du BBT-DARIAH pour soulever différents points et notamment celui des entités surnaturelles. La discussion est encore en cours du fait de l'absence de consensus autour des suggestions faites. Un point de discussion a notamment émergé au sujet de l'existence de catégories de « personnages littéraires » déjà existante sous la facette Objets propositionnels, tandis que les types de personnages étaient classés sous Rôles.

Dans le cadre du projet, les *collections* ont malgré tout été déplacées selon ce qui avait été envisagé en accord avec PACTOLS dans le micro-thésaurus Objets Propositionnels sous Objets Conceptuels (ce terme classe les concepts mentaux produits par l'activité humaine, représentant des données de réalités physiques, logiques ou psychologiques, ou même de fiction ; il s'agit d'une terminologie établie par DARIAH). Le terme de classement unique « religion et mythologie » envisagé en réunion avec le comité scientifique a été repris pour servir de chapeau aux puissances suprahumaines et figures.

---

<sup>7</sup> À ce propos, voir la présentation de Blandine Nouvel, Evelyne Sinigaglia, Véronique Humbert : « Deconstructing for reconstructing: the use of the Backbone Thesaurus for reorganising the PACTOLS Thesaurus » durant la conférence FAIR Heritage: DigitalMethods, Scholarly Editing, and Tools for Cultural and Natural Heritage (17-18 juin 2020) retranscrite ici ; <https://halshs.archives-ouvertes.fr/halshs-02890892>

## 5.1.2 Préparatifs et projections pour le déplacement de branches

Aussi, et en toute fin de projet, il a été décidé de déplacer de manière préventive l'ensemble des entités traitées dans la branche Objets Propositionnels de PACTOLS, quitte à ce que ces classifications soient revues par la suite. Par précaution, les manipulations ont d'abord été réalisées dans un second thésaurus d'exercice avant d'être reconduites dans le thésaurus de PACTOLS. La première étape a été de renommer les termes de regroupement (divinité, etc.) selon le vocabulaire choisi (puissances suprahumaines), dans la stricte limite de notre corpus. Ces termes de regroupement ont eux-mêmes été multipliés et reclassés pour correspondre à l'arborescence des aires civilisationnelles élaborée avec le comité scientifique. Seuls les termes sous lesquels des entités venaient à être classées ont été créés. Enfin, les branches *figures* (héros) et *puissances suprahumaines* (divinités) ont été toutes deux déplacées sous le terme *religion et mythologie* créé à l'occasion sous Objets propositionnels. La manipulation nécessaire au déplacement de ces branches est extrêmement simple sous OpenTheso, où il suffit de sélectionner le terme générique ou spécifique et de choisir d'ajouter une nouvelle relation spécifique ou générique, puis de supprimer l'ancienne relation chez le terme spécifique.

Il faut noter plusieurs éléments particuliers. Tout d'abord, le terme *panthéon grec* existant dans l'ancien micro-thésaurus *Sujets* a été ajouté sous *religion et mythologie* (sous un terme provisoire de *regroupement de divinités*). Cela a permis de déplacer le terme *Olympiens*, anciennement terme de classement sous *divinités grecques* sous *panthéon grec*, pour aller dans le sens de l'opinion du comité scientifique considérant ce terme comme un regroupement historiographique plutôt qu'un ensemble de classement pertinent de puissances suprahumaines.

D'autre part, d'autres entités telles que *cycle troyen* ou *conquêtes de Zeus* ont elles aussi été déplacées dans un autre sous-ensemble d'Objets propositionnels : *mythes grecs*. Leurs liens de terme générique à spécifique avec les figures qu'ils classaient ont été modifiés en liens d'association pour ne pas perdre la trace de ces précédentes connexions.

## 5.2 Chantiers de réindexation

Les notices d'autorité servent à décrire des entités et accéder à une documentation. Il faut donc prendre en compte la nature des entités du corpus, particulièrement pour le cas des notices qui, avant le travail effectué, pouvaient confondre plusieurs entités, ou bien dont les noms étaient susceptibles d'entraîner une confusion avec d'autres entités sans rapport du fait d'une pauvreté de la notice.

Il s'agit d'une mission encore différente, qui nécessite un suivi pour permettre la transmission des éléments à un prochain travail de nettoyage et de recatalogage, à envisager potentiellement dans un cadre plus large d'une normalisation des choix des termes d'indexations à retenir pour la description des notices. Dans IdRef, un outil comme [paprika.idref.fr](http://paprika.idref.fr) peut aider à rapprocher notices bibliographiques et notices d'autorité.

Il n'est évidemment pas possible de dresser une liste exhaustive des indexations erronées ou abusives. Plusieurs pistes sont à privilégier :

- les notices désambiguïsées : deux voire plusieurs notices ont été créées à partir d'une seule, lorsque celles-ci invitaient à la confusion entre des figures ou puissances distinctes. Une notice préexistante a été maintenue, conservant généralement l'identité

la plus à même à avoir une bibliographie plus conséquente. Celles créées à sa suite n'indexent pour l'instant aucune notice supposée les concerner.

- les noms communs : certaines entités, par leur nom même et malgré les indices devant permettre de les identifier, indexent des homonymes malheureux, noms de personne ou noms communs.

Plusieurs notices ont ainsi pu être désambiguïsées durant la réindexation comme par exemple :

Créon (mythologie grecque), divisé en Créon de Thèbes / Créon de Corinthe.

Diomède (mythologie grecque), divisé en Diomède roi de Thrace / Diomède roi d'Argos.

Inanna (divinité sumérienne), divisé en Inanna / Nin-Egala.

Ashtart (divinité sémitique), divisé en Astarté / Ishtar / Ashtar / Shaushka / Atargatis / Baalat Gubal.

Le Sudoc étant un outil collectif et collaboratif, des modifications pourront être effectuées malgré le marquage en zone 686 (\$aRefDivinités\$c Appel à projets 2019\$2 Collex-Persée). Nous avons ainsi pu faire l'expérience d'une confusion sur la notice « Hermès (divinité grecque) ». Un autre projet CollEx-Persée ([Mistara de la BULAC](#)) l'a fusionnée avec la notice « Hermès Trismégiste (auteur prétendu ; divinité grecque) », qui faisait pourtant référence à une entité bien spécifique, issue d'un syncrétisme avec le dieu égyptien Thot, et réputée avoir écrit le corpus hermétique. À la suite de cette intervention, un travail de désambiguïsation a été nécessaire et une ré-indexation des documents concernés est en cours. Sans remettre en question le travail collaboratif dans le Sudoc et les nécessaires améliorations qui peuvent être apportées par d'autres, il apparaît néanmoins nécessaire de protéger les travaux de référencement scientifique réalisés dans le cadre de projets CollEx.

### **5.3 Quel statut pour ces notices s'il ne s'agit plus de personnes avec la transition bibliographique ?**

Depuis novembre 2015, les bibliothèques françaises sont engagées, sous l'égide de la BnF et de l'Abes, dans un processus de transition bibliographique, qui vise à exposer les catalogues dans le web de données. Ce programme donne lieu à la rédaction d'un nouveau code de catalogage : inspiré de modèles conceptuels préexistants (FRBR, puis IFLA-LRM), RDA-FR devrait permettre de distinguer des entités (œuvre, expression, manifestation, item) actuellement enchevêtrées dans les notices bibliographiques.

À terme, IdRef sera forcément bouleversé par ce nouveau modèle. L'Abes et la BnF travaillent en ce moment à un fichier national d'entités (FNE), qui incorporera, au-delà des notices d'autorité traditionnelles, la plupart des entités définies par le modèle IFLA-LRM (agents, œuvres, concepts, laps de temps, lieux). Un premier déploiement du FNE est prévu en 2023.

D'ores et déjà, il semble assuré que les divinités et héros ne seront plus considérés comme des personnes dans le nouveau modèle. Publiée en décembre 2019, la section 3 "Enregistrement des attributs des agents", considère en effet comme relevant de la catégorie personne "un agent identifié comme étant un individu existant ou ayant existé, ou une entité supposée avoir une responsabilité intellectuelle sur une ressource (par exemple : un esprit, un animal, etc.)" (p. 3). On voit que n'entreront dans ce cadre strict que quelques "curiosités" de

notre corpus, comme Hermès Trismégiste. Mais on peut se poser la question de l'extension de cette responsabilité, si celle-ci peut prendre la forme d'adresse ou dédicace dans le cadre d'hymnes, poèmes et documents épigraphiques.

Presque tous les héros et divinités risquent de passer dans la catégorie des "concepts". Ceux-ci n'ont pas vocation à être décrits dans le code de catalogage RDA-FR. À l'heure actuelle, ils dépendent de la réflexion du sous-groupe de travail "Concepts, lieux, temps", créé en juin 2017, dirigé par Florence Ménard, positionné entre le programme de transition bibliographique et le Centre national Rameau. En pratique, cela signifie que, dans les années à venir, divinités et héros risquent d'intégrer le vocabulaire Rameau. Ce qui déplacera certaines difficultés : la question ne sera plus, par exemple, d'arriver à distinguer divinités plurielles (noms de personnes) et catégories de personnes mythologiques (noms communs), mais plutôt individus ayant réellement existé ou non (on entrevoit déjà une zone grise, constituée des personnages historiques divinisés, des figures légendaires, etc.).

Une distinction claire existe parfois sans être systématisée, dans IdRef mais aussi à l'origine dans des dictionnaires et lexika, entre une figure historique et son alternative mythologique ou légendaire. Souvent, cette division est due à l'existence d'appellations différentes entre la figure historique et son miroir, mais aussi à la confusion possible entre différents référents réels : Sardanapale (nom donné par Diodore) est souvent associé à Assurbanipal, mais son personnage s'inspire de plusieurs autres rois d'Assyrie. On peut donc imaginer une distance prise entre les figures historiques (agents potentiels de textes, tablettes, documents administratifs, lettres), et ces émanations (concepts) qui serviront plus directement d'inspiration aux récits et à l'iconographie. Cet enjeu est pour l'instant évoqué pour les personnages légendaires du guide du catalogueur sans toucher le cas de la mythologie (alors que la mythologie romaine était conçue comme historique). On peut se demander si la distinction continuera de suivre les mêmes règles. Pour tout personnage historique qui déborde sur le mythe, dont le nom ne varie pas et qui n'a pas d'autre référent historique, la caractérisation devrait clairement s'établir du côté de l'agent, même si elle peut garder un qualificatif "mythologie romaine" et avoir plusieurs termes génériques renvoyant à son caractère réel (rois de Rome / mythologie romaine pour Romulus par exemple)

À partir du moment où les divinités et héros ne relèveront plus des notices d'autorité "personnes", ils ne seront *a priori* plus créés et modifiés librement par les catalogueurs du réseau Sudoc, mais seront gérés par la BnF. Il paraît donc indispensable que les conclusions de notre travail soient transmises, non seulement à l'Abes, mais aussi et surtout au Centre national Rameau. Dans un sous-domaine bien identifié, les spécialistes de la BnF auront ainsi à disposition un référentiel riche, structuré et hiérarchisé de façon rigoureuse, validé par des chercheurs, aligné tant avec PACTOLS qu'avec Wikidata.

## **5.4 Enrichissements complémentaires : bases de données fermées et formes d'entités particulières**

Le projet RefDivinités avait envisagé au départ un travail d'alignement avec le LIMC-France qui a produit plusieurs bases de connaissances partant du travail encyclopédique du *Lexicon Iconographicum Mythologiae Classicae*. Cependant leur base de données propriétaire présente des difficultés de taille. La recherche sur le site [limcfrance](http://limcfrance.fr) est divisée en trois sections : LIMC-icon pour accéder aux documents iconographiques, LIMC-biblio pour accéder à la documentation qui a pour bibliographie des articles ou éléments iconographiques du LIMC, et enfin LIMC-abrev, pour retrouver les articles du LIMC à partir des abréviations en usage. Parmi ces trois portails de recherche, seul le LIMC-icon donne accès à une notice d'entité possédant une URL propre : "www.limc-france.fr/objet/+numéro de l'objet".

Un vocabulaire contrôlé structure la description des objets : le thésaurus ThéA, pensé depuis 1981 par l'équipe du LIMC, devenu multilingue en 1998 est en processus de normalisation pour correspondre aux normes ISO 25694 de thésaurus et aux recommandations SKOS. Les noms de figures et puissances figurent dans ce thésaurus. Malgré l'ouverture sur le web sémantique envisagée par le portail CLAROS<sup>8</sup>, il n'y a pas aujourd'hui de passerelle permettant de lier le vocabulaire iconographique du LIMC à d'autres plateformes.

De manière sensiblement différente, il n'est pas envisageable pour le moment de réaliser d'alignement avec l'*Année philologique*. Le travail et la richesse du vocabulaire d'indexation sont à signaler, mais il faut d'une part souligner l'absence de normalisation de ces termes, ainsi que le nombre potentiellement important de doublons. Surtout, ils ne possèdent pas d'identifiant pérenne. Enfin, la base est elle aussi fermée.

## 5.5 Communications autour du projet

La mission, qui devait s'achever en avril 2020, a été ralentie comme tant d'autres par la pandémie de COVID-19. La journée d'étude prévue au sein de la BIS pour clôturer ce projet n'a malheureusement pas pu avoir lieu. Des communications internes et via les réseaux de la Bibliothèque ont toutefois pu être diffusées.

L'implication de l'Abes a permis la publication d'un premier billet dans Fil'Abes en septembre 2019 : <https://fil.abes.fr/2019/09/23/projet-collex-persee-refdivinites-rapprocher-idref-et-pactols/>. Un second billet a été publié en novembre 2020 sur le blog technique de l'Agence, Punktokomo : <https://punktokomo.abes.fr/2020/11/19/divinites-et-heros-du-monde-mediterraneen-antique-retour-sur-le-projet-refdivinites/>.

Le projet RefDivinités a été présenté par Sébastien Dalmon au [séminaire des référents CollEx du 13 novembre 2019](#), dans le cadre d'une intervention plus globale menée à deux voix avec Amélia Laurenceau, cheffe du Service de la Valorisation numérique des collections et du soutien à la recherche, sur l'animation de réseaux et le portage de projets dans le cadre des délégations CollEx-Persée de la BIS.

Le fonctionnement particulier du comité scientifique a par ailleurs fait l'objet d'un focus (p. 17) dans le [rapport d'activité CollEx-Persée de 2019](#).

Le projet RefDivinités a donné lieu à un poster au colloque [« Naming and Mapping the Gods in the Ancient Mediterranean. Spaces, Mobilities, Imaginaries »](#), organisé par Corinne Bonnet et son équipe du programme ERC Mapping Ancient Polytheisms (Université Toulouse Jean Jaurès). Ce colloque devait se tenir à l'origine du 25 au 27 mars 2020, mais a été repoussé du 10 au 12 février 2021. En raison des contraintes sanitaires, il a eu lieu uniquement en distanciel sur zoom. Une présentation bilan du projet a ainsi pu être réalisée en ligne par Alyx Taounza-Jeminet et Sébastien Dalmon le 12 février 2021, ce qui a permis de le faire connaître à des chercheurs à l'échelle internationale. Le poster de présentation est disponible en anglais et en français sur le site CollEx-Persée.

---

<sup>8</sup> L'adresse <http://www.clarosnet.org> est aujourd'hui inaccessible. Pour plus d'informations sur le LIMC et CLAROS, voir Anne-Violaine Szabados, « Du système documentaire du LIMC au portail CLAROS : Interopérabilité et optimisation de l'information archéologique grâce à l'usage de normes », JIAP 2010 : <https://halshs.archives-ouvertes.fr/halshs-00524724>

## Conclusion

Le projet Collex-Persée RefDivinités est avant tout une expérimentation : dans la transition vers un nouveau schéma de thésaurus, dans la redéfinition d'une terminologie, dans le dialogue entre les métiers de la conservation, des bibliothèques et du numérique avec ceux de la recherche. La Bibliothèque interuniversitaire de la Sorbonne, associée à Frantiq, a pu mener à bien cette mission, et ouvrir dans le même temps de nouvelles portes. Certaines concernent directement le corpus en jeu, sa ré-indexation, sa classification qui se veut plus pérenne dans le temps. D'autres invitent à réfléchir d'une part aux outils, les usages d'OpenTheso par exemple, mais aussi aux bouleversements qui touchent à la structure et la description même des données au sein de la transition bibliographique, pour une meilleure stabilité et interopérabilité à grande échelle, au-delà des ponts essentiels entre IdRef et PACTOLS.

Le projet RefDivinités ne saurait être considéré comme totalement achevé. De nouvelles entités seront probablement à créer par la suite, au gré de nouvelles publications et de nouveaux besoins bibliographiques. Le travail pourra être également poursuivi, par d'autres acteurs, pour les puissances suprahumaines et figures d'autres périodes et d'autres aires civilisationnelles. Ce projet d'alignements entre autorités se veut répliquable à d'autres domaines. Ainsi, une documentation a été fournie pour servir au projet Collex-Persée [ArchéoAI](#) (enrichissement des autorités IdRef et PACTOLS de sites archéologiques et alignements de référentiels noms de lieux et géolocalisation).

### **Quelques chiffres :**

664 entités ont été alignées entre PACTOLS et IdRef. À celles-ci s'ajoutent dans notre tableau final 7 autorités RAMEAU considérées comme des noms communs, qui n'ont pas pu être traitées dans IdRef mais vers lesquelles renvoient les correspondants exacts dans PACTOLS (Courètes, Dieux de Samothrace, Hyperboréens, Lamassu, Lares, Muses, Pénates). D'autres noms communs (amazones, centaures, sirènes par exemple) n'ont pas été alignés, l'autorité Rameau dépassant le seul cadre du monde méditerranéen antique.

431 entités existaient dans PACTOLS avant le début du chantier, dont plusieurs ont été reclassées (iconographie, concepts historiographiques, abstractions, classements sans pertinence). L'arborescence « religion et mythologie » comporte aujourd'hui 802 entités (dont plusieurs hors corpus et de classement).

392 entités existaient dans IdRef (dont 248 exportées de la BnF). 272 ont été créées.

434 alignements ont été faits vers data.bnf.fr ; 563 vers le VIAF ; 647 vers Wikidata.



# Annexes

## Annexe 1 : Liste des membres du conseil scientifique

Ces personnes, aux spécialités variées, ont pu apporter leur expertise ponctuelle ou régulière lors de réunions et entretiens, et ont marqué leur intérêt pour ce travail documentaire au service de la recherche. Ils ont été réunis par Sébastien Dalmon.

[Azarnouche Samra](#), EPHE

[Belayche Nicole](#), EPHE

[Bonnet Corinne](#), Université Toulouse Jean Jaurès

[Briquel Dominique](#), Sorbonne Université

[Bruit-Zaidman Louise](#), Université de Paris

[Delattre Charles](#), Université de Lille

[Ellinger Pierre](#), Université de Paris

[Fenet Annick](#), ENS

[Georgoudi Stella](#), EPHE

[Glassner Jean-Jacques](#), EHESS et CNRS

[Golosetti Raphaël](#), Sorbonne Université

[Hawley Robert](#), EPHE

[Lacam Jean-Claude](#), Université Paris 1 Panthéon-Sorbonne

[Mouton Alice](#), CNRS

[Pirrenne-Delforge Vinciane](#), Collège de France

[Rendu-Loisel Anne-Caroline](#), Université de Strasbourg

[Sebai Meriem](#), Université Paris 1 Panthéon-Sorbonne

[Tallet Gaëlle](#), Université de Limoges

[Traina Giusto](#), Sorbonne Université

[Wylter Stéphanie](#), Université de Paris

## Annexe 2 : Diagramme des « aires religieuses »

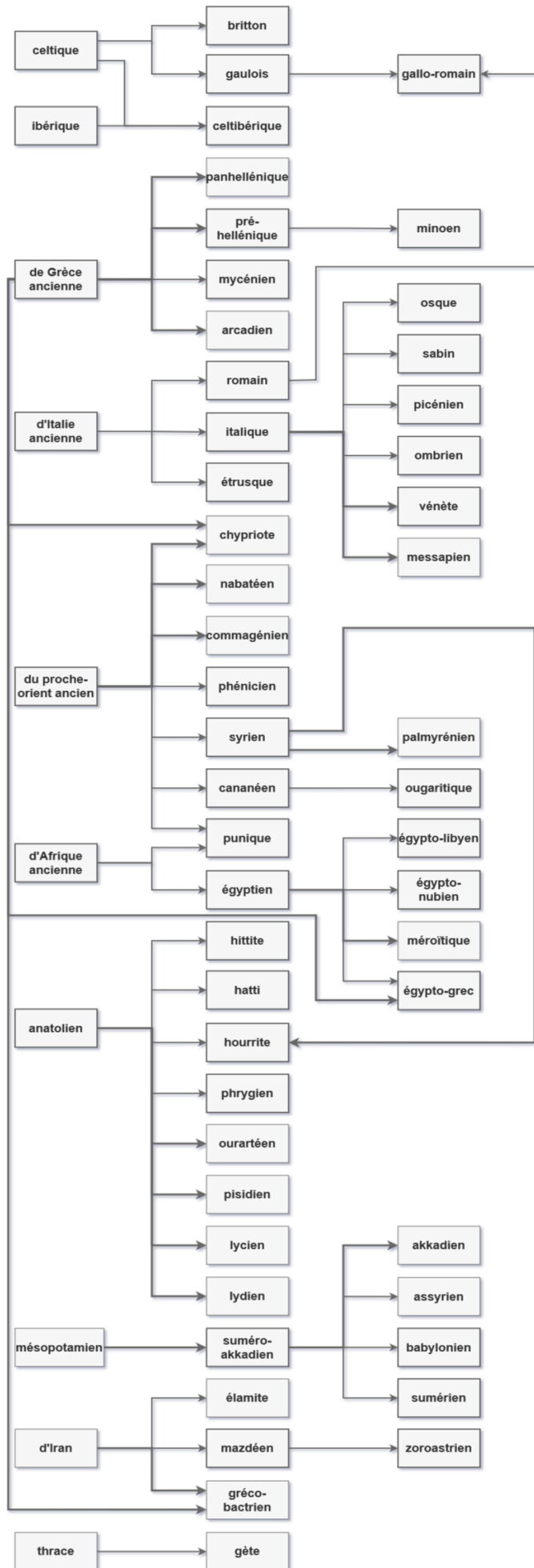
Ce diagramme a été conçu en collaboration avec le conseil scientifique. Il est le fruit de suggestions, contradictions et accords qui ont rythmé les débats lors de réunions et échanges. Son application partielle a été faite dans le thésaurus PACTOLS. Elle n'a pas en soi de valeur normative, et peut évidemment évoluer suite à des remarques, ajouts, actualisations au sein du réseau Frantiq. Elle doit avant tout servir les catalogueurs dans leur travail comme les étudiants et scientifiques dans leur recherche.

Cette arborescence **n'a pas vocation à indexer des peuples ou des langues**. Nous sommes dans un entre-deux visant à l'indexation de puissances suprahumaines. Celles-ci prennent dans certains cas part à une religion déterminée (*mazdéisme*) ; ailleurs, elles sont rattachées à une culture, voire un lieu. Pour cette raison, certains termes ont pu être supprimés comme ne servant pas à la description d'entités particulières (*égéen, chypriote*).

D'autre part, nous avons pu retenir des termes à valeur **géographique** faisant sens (*d'Italie ancienne*, etc.) et nous avons tenté d'évacuer ou de moduler des termes qui apparaissaient relever plus exclusivement de la **linguistique** (*sémitique, osque* etc.).

Cette arborescence souhaite représenter autant que faire ce peu **une vision actuelle de la recherche**. Des thèmes obsolètes selon l'historiographie la plus récente ont été évacués (*assyro-babylonien, oriental*, etc.) pour laisser la place à d'autres propositions.

Cette arborescence, enfin, n'est pas isolée du reste du thésaurus PACTOLS. Il faut envisager des liens d'associations entre ces termes et ceux plus strictement géographiques ou de populations, mais aussi potentiellement avec les termes historiographiques (*oriental*) indexés en un autre emplacement et dont on souhaiterait restreindre l'usage.



## Annexe 3 : Arborescences et terminologies dans PACTOLS et IdRef en avril 2019

Les termes retenus sont ceux du corpus de RefDivinités.

### IdRef

Les deux arborescences suivantes, « Religions » et « Dieux » présentent des noms communs du vocabulaire Rameau et à ce titre non modifiables dans IdRef. Les notices de religions renvoyant aux notices de mythologies et dieux affiliées qui contiennent elles-mêmes les figures et puissances du corpus.

- Religions
  - celtique
    - britton
    - gaulois
      - gallo-romain
  - égyptien
  - germanique
  - ibérique
  - gréco-romain
    - grec
      - mycénien
      - chypriote
      - de Samothrace
    - italique
      - romain
      - étrusque
      - vénète
  - pré-hellénique
    - égéen
    - minoen
  - thraco-phrygien
    - phrygien
  - Oriental
    - d'Asie mineure
    - syrien
    - sumérien
    - assyro-babylonien
      - assyrien
      - babylonien
    - cananéen
    - hittite
    - sémitique
      - ougaritique
      - phénico-punique
        - phénicien
        - punique
    - mazdéen
      - zoroastrien

- Dieux
  - Anatoliens
  - Assyro-babyloniens
  - Cananéens
  - Celtiques
    - Gaulois
  - Egyptiens
  - Etrusques
  - Grecs
    - Minois
    - De Samothrace
  - Hittites
  - D'Ougarit/ ougaritiques
  - Phéniciens
  - Phrygiens
  - Punique
  - Romains
  - Sumériens / mésopotamiens
  - Syriens
  - Védiques
  - Zoroastriens

#### Qualificatifs :

Utilisés dans les noms d'entités idRef pour éviter les confusions sous la forme « Puissance (divinité(s) + qualificatif) » ou « Figure (mythologie + qualificatif) »

- assyro-babylonienne
- babylonienne
- cananéenne
- celtique
- d'Asie mineure
- dace
- égyptienne
- étrusque
- gallo-romaine
- gauloise
- grecque
- hittite
- italique
- mycénienne
- orientale
- phénicienne
- phrygienne
- romaine
- sémitique
- sumérienne
- syrienne
- thraco-phrygienne
- védique

- vénète
- zoroastrienne

## **PACTOLS**

Le corpus de RefDivinités s'est consacré aux entités dans la branche Anthroponymes. Mais les champs de la religion et de la mythologie antique sont aussi contenus dans la branche Sujets. Cette arborescence est actuellement en cours de restructuration.

Anthroponymes

Divinités

Abstractions divinisées

Egyptiennes

Etrusques

Gauloises

Germaniques

Grecques

Hittites

Mazdéennes

Orientales

Phéniciennes

Romaines

Héros

Gaulois

Grecs

Orientaux

Romains

Sujet

Religion

Divinités

Panthéon égyptien

Panthéon étrusque

Panthéon gaulois

Panthéon germanique

Panthéon gréco-romain

    Panthéon grec

    Panthéon romain

Panthéon oriental

Panthéon phénico-punique

    Panthéon phénicien

    Panthéon punique

Mythologie

    Mythologie égyptienne

    Mythologie gréco-romaine

        Mythologie grecque

        Mythologie romaine

    Mythologie orientale

        Mythologie assyro-babylonienne

        Mythologie ougaritique

        Mythologie sémitique

Religions

    Hermétisme

    Manichéisme

    Mazdéisme

    Religion celte

        Religion gauloise

        Religion ibérique

    Religion égyptienne

Religion germanique

Religion gréco-romaine

Religion grecque

Religion chypriote

Religion mycénienne

Religions italiques

Religion étrusque

Religion romaine

Religion orientale

Culte isiaque

Mithraïsme

Religion assyro-babylonienne

Religion hittite

Religion sémitique

Religion phénicienne

Religion punique

Religion sumérienne

Zoroastrisme

Religion védique

Religions pré-helléniques

Religion égéenne

Religion minoenne