

AAP



**CollEx-Persée**  
2021\_2022

Le Rapport de synthèse

**Conseil scientifique du GIS**  
**du 20/10/2022**



**A**u terme d'un appel à projets ouvert le 15 juin 2021 et clos le 10 décembre 2021, 31 dossiers ont été déposés auprès du GIS CollEx-Persée. Une évaluation en plusieurs étapes a été organisée : expertises des professionnels de l'information scientifique du 20 décembre 2021 au 18 février 2022 et évaluation par des binômes du conseil scientifique du 28 février au 28 avril 2022.

A l'issue du processus, le conseil scientifique du GIS a désigné, en séance plénière du 18 mai 2022, les lauréats. Les résultats ont été présentés au conseil des membres du GIS, le 30 mai 2022.

Ils sont à retrouver sur le web : [https://www.collexpersee.eu/aap21\\_22-resultats/](https://www.collexpersee.eu/aap21_22-resultats/)

### **19 projets ont été financés à hauteur de 1.679.019 euros.**

Le GIS CollEx-Persée, en tant qu'infrastructure de recherche en information scientifique a créé ce mécanisme d'appel à projets pour favoriser le développement de services à la recherche adossés aux collections d'excellence au sein du réseau et encourager des démarches collaboratives de numérisation, d'enrichissement de métadonnées et/ou de développement d'autres types de services sur des corpus de différentes natures.

Il s'agit du troisième appel à projets collaboratifs depuis 2018. Pour rappel, CollEx-Persée a financé au total **92 projets pour près de 5,6 millions d'euros** :

- en 2018, 41 projets (sur 93 déposés, en deux vagues) ont été retenus pour la somme de 1.588.670,77 euros.
- en 2020, 25 projets (sur 46 déposés) pour la somme de 1.997.886 euros.
- En 2022, donc, 19 projets (sur 31 déposés) pour la somme de 1.679.019 euros.
- A cela s'ajoute l'appel à résidences chercheurs, qui a financé 7 contrats annuels, en 2021-2022 à hauteur de 299.177 euros.

Ce troisième et dernier appel à projets collaboratifs du GIS CollEx-Persée devait **consolider les acquis du mécanisme** pour susciter des projets numériques à valeur ajoutée, favoriser la dynamique de co-construction entre chercheurs et professionnels de la documentation, faire des bibliothèques des acteurs de la recherche au plus près des besoins des chercheurs, articuler les projets aux collections patrimoniales et scientifiques, numériques et matérielles des bibliothèques de recherche, encourager le fonctionnement en réseau, s'inscrire dans les objectifs de la science ouverte, s'ouvrir à toutes les disciplines et à l'interdisciplinarité.

Compte tenu de l'expérience acquise dans les éditions précédentes, **certain éléments ont été particulièrement renforcés et évalués en conséquence par les experts** : la mise en place d'un réseau d'acteurs complémentaires au-delà des institutions porteuses ; la qualité de l'ingénierie de projet mise en place en termes de gouvernance de projet ou de pérennisation des livrables ; la présence d'un volet formation et valorisation pour une meilleure diffusion des résultats ; l'instruction des questions juridiques et ayants droit ; l'inscription dans les principes de la science ouverte par la mise en place d'un PGD et la préservation numérique et, enfin, l'ouverture à l'international.

Il a été envisagé d'aller plus loin, au moment de la conception de l'appel, et de demander aux porteurs de démontrer l'intérêt du service proposé à l'échelle nationale ou d'un secteur disciplinaire ou thématique ; d'établir l'aspect mutualisable du ou des services proposés ; de proposer une gouvernance et un modèle économique viables au-delà de la durée du projet lui-même permettant d'assurer la pérennité et l'enrichissement du service proposé. C'était une approche sans doute trop ambitieuse, nécessitant des financements pluriannuels. Cette démarche a été réservée pour nourrir la réflexion autour d'un CollEx-Persée n°2 (2023-2027) qui devrait s'organiser de manière programmatique.

Le conseil scientifique du GIS CollEx-Persée a noté favorablement la qualité générale des projets déposés. Il est rare qu'un projet ne soit pas au niveau d'un point de vue technique. Un nombre croissant de déposants prend en compte la pérennité de la proposition et intègre des développements pour servir à une communauté élargie. Une vraie réflexion préside à la démarche scientifique. Une attention accrue dans la mise en place des partenariats est à souligner.

Seuls, 5 à 6 projets ont semblé bien trop limités dans leur propos général, tandis que quelques autres ont semblé trop fragiles sur certains aspects, soit scientifiques, soit de gouvernance, soit dans le montage financier, par exemple.

L'ensemble des avis des experts sont transmis aux porteurs dans un souci de transparence et de montée en compétences.

Cette expérience est importante pour l'ensemble de la communauté scientifique et de l'information scientifique et technique. Elle a agi comme un levier de transformation des pratiques inaugurant un véritable rapprochement entre chercheurs, bibliothécaires, opérateurs, faisant de ces deux derniers de véritables acteurs des projets de recherche.

Il conviendrait, désormais, de cartographier finement l'ensemble des projets financés par le GIS en vue de démontrer le caractère unique de ce dispositif, complémentaire d'autres mécanismes de financement comme l'ANR.

Il serait aussi souhaitable de pouvoir réunir cette communauté, pour échanger et capitaliser sur les résultats afin de déterminer des axes communs susceptibles de revêtir une dimension nationale soutenus par le dispositif dans son nouveau projet 2023-2027.

→ Tous les détails sont à retrouver sur le site web CollEx-Persée :

<https://www.collexpersee.eu/les-projets/>

→ Les attendus et critères de l'appel à projets sont à retrouver :

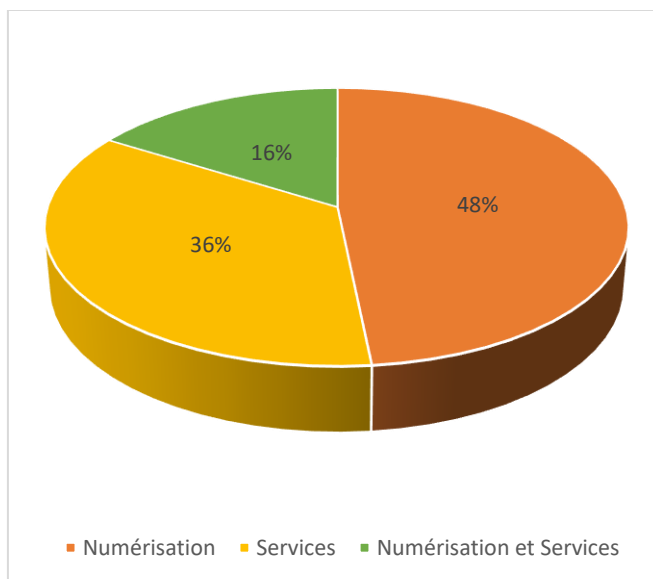
[https://www.collexpersee.eu/appel-a-projets-2021\\_2022/](https://www.collexpersee.eu/appel-a-projets-2021_2022/)

Rédaction : [catherine.desos-warnier@bnu.fr](mailto:catherine.desos-warnier@bnu.fr)

## Table des matières

Statistiques de dépôts et typologie des projets.....	5
Porteurs et partenaires engagés dans les projets.....	6
Actions à réaliser dans le cadre des projets.....	7
Disciplines représentées au sein des projets déposés.....	8
Annexe 1 : Lauréats de l'appel à projets CollEx-Persee 2021-2022.....	9
Annexe 2 : Liste des experts en IST sollicités.....	11
Annexe 3 : Résumés en français et en anglais des projets retenus.....	12
<b>BE_NUM</b> .....	12
<b>CHI-KNOW-PO</b> .....	13
<b>CHRONOSPEDIA</b> .....	14
<b>CoBaLPP</b> .....	15
<b>CoESciTer</b> .....	16
<b>COREL</b> .....	17
<b>Corpus SATHMA</b> .....	18
<b>DreAM</b> .....	19
<b>ICB</b> .....	20
<b>IndexAngkor</b> .....	21
<b>IRIS</b> .....	22
<b>ISSA2</b> .....	23
<b>NumeNord</b> .....	24
<b>NUMISMATES</b> .....	25
<b>ORDY GOAL</b> .....	26
<b>ORESM</b> .....	27
<b>PRET19</b> .....	28
<b>Rev@ntiq</b> .....	29
<b>TariMa</b> .....	30

## Statistiques de dépôts et typologie des projets



### **31 projets déposés (dont 1 inéligible)**

Coût total : 6 552 355,45 euros €

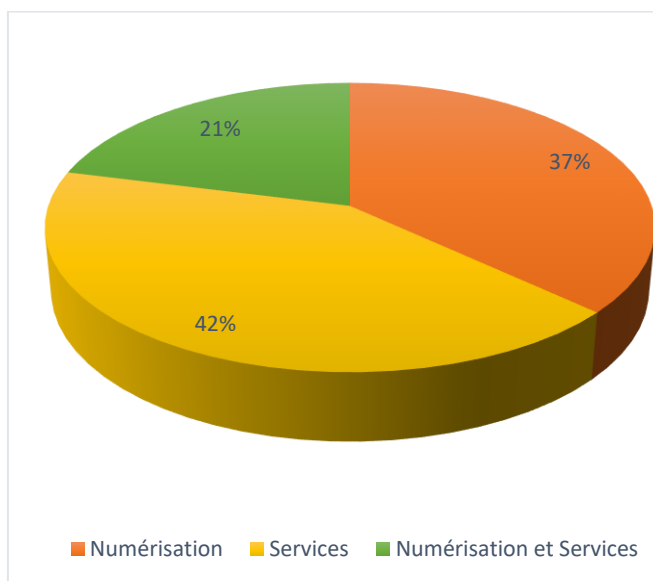
Demande de subvention CollEx-Persée :  
2 725 617,63 euros € = 42 % du coût total

Services : 11

Numérisation : 15

Numérisation et Services : 5

- **Remarque** : 6 projets déposés sont le volet n° 2 d'un projet déjà financé par CollEx-Persée : 3 sont lauréats.
- **Remarque** : 3 projets, non retenus dans l'AAP 2019 se présentent à nouveau en ayant tenu compte des avis des experts : ils sont lauréats.



### **25 projets lauréats**

Coût total : 4 159 363,62 €

Subvention CollEx-Persée : 1 679 019,63 €  
= 40 % du coût total

Numérisation : 7

Numérisation et Services : 4

Services : 8

**Remarque 1** : CollEx-Persée soutient financièrement à hauteur de 40% des projets co-construits IST/recherche, demeurant ainsi un acteur essentiel du soutien à la recherche française, inscrit dans la science ouverte.

**Remarque 2** : On ne peut que constater un essoufflement dans la dynamique de dépôts (31 projets contre 46 déposés en 2019), mais avec un très bon niveau qualitatif globalement.

**Remarque 3** : les porteurs administratif restent des membres du réseau : 8 établissements délégataires ou associés pour 11 projets déposés (BIS/UParis 1, BULAC, BNU, INHA, DBU/Paris 3, Bordeaux Montaigne, ULille et UGA) et 12 établissements labellisés CollEx pour 15 projets déposés (EFEO et EFR, GDS Frantiq, CEALex, Collège de France, CNAM, CIRAD, CRBC, MOM, MSH de Poitiers, Clermont, Bourgogne). Seuls 5 porteurs sont plus à la marge du système CollEx-Persée (Ecole d'économie de Paris (PSE : Paris school of economics), U Polytechnique des Hauts de France, EPHE-PSL et Institut de mécanique céleste et de calculs des éphémérides de l'Observatoire de Paris et MSH de Nantes).

## Porteurs et partenaires engagés dans les projets

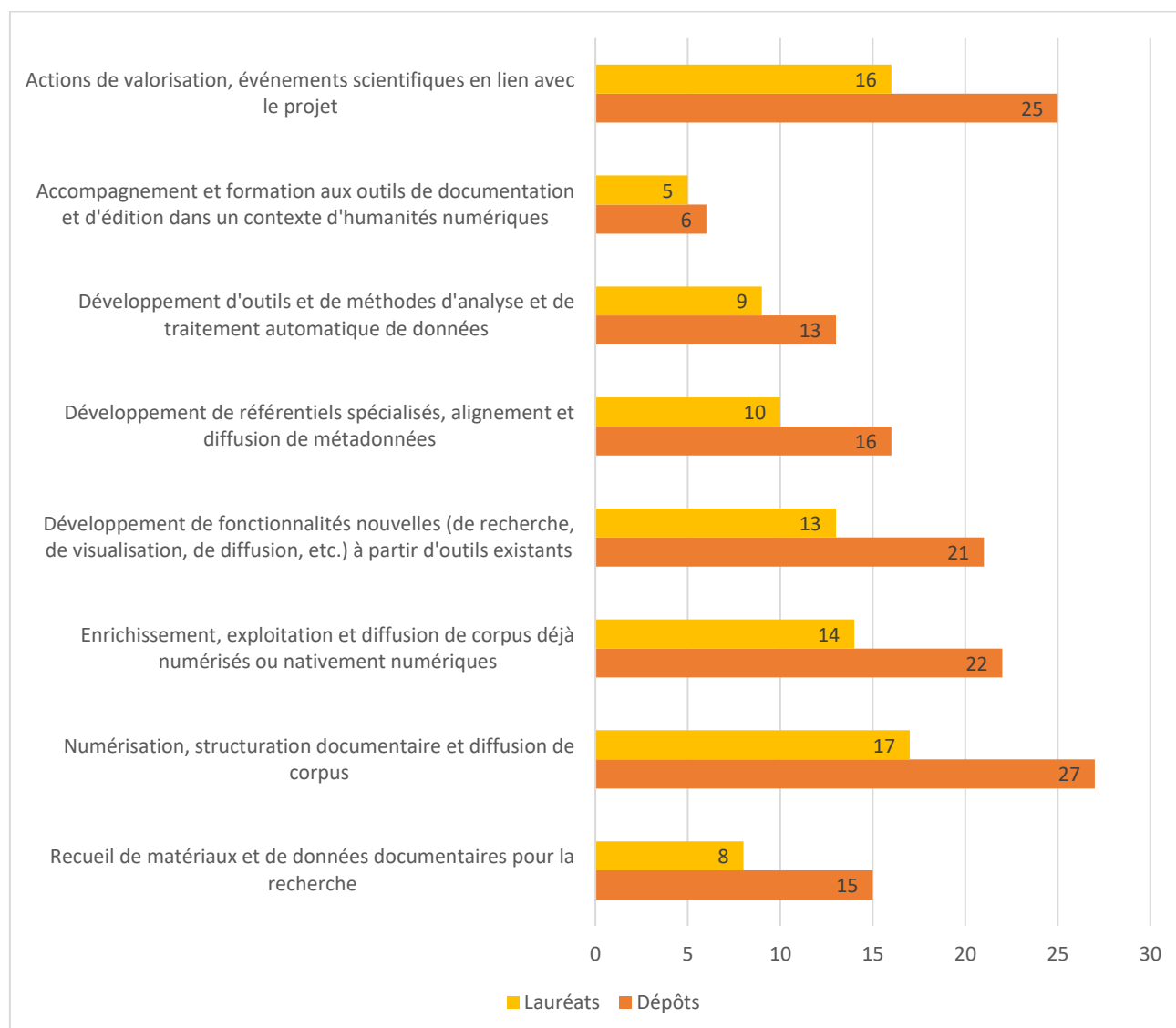
174 partenaires engagés dans les projets, dont 24 deux fois ou davantage.

- **Opérateurs CollEx** : 4 opérateurs dans 7 projets (*Persée, INIST, Abes dans 2 projets, BnF dans 3*).
- **SCD ou BU** : 17 établissements dans 15 projets (*dont BULAC dans 3 projets, BIS dans 4, BNU et BSG dans 2 projets chacune, SCD Bordeaux dans 3*).
- **Ecoles françaises à l'étranger** : 4 EFE dans 5 projets (*EFA, EFR, EFEO dans 3 projets, IFAO dans 2*).
- **Instituts, Ecoles, Muséum, Observatoire, Collège de France** : 17 institutions dans 14 projets (*Collège de France dans 3 projets, EPHE-PSL dans 2 projets*).
- **Unité propre de recherche (UPR)** : 3 dans 2 projets (*IRHT, Centre de Recherches sur les Littératures et la Sociopoétique, Centre d'histoire Espaces Cultures*).
- **Institut d'études avancées (IEA)** : 1 (de Strasbourg) dans 1 projet.
- **Unité mixte de recherche (UMR)** : 34 dans 18 projets (*et Archéorient, ArAr, HISOMA, LAHRA, Archimède, URMIS, ASM dans 2 projets chacune*).
- **Fédération de recherche (FR)** : 1 dans 4 projets : la MOM.
- **Unité mixte de services (UMS)** : 1, le CRBC et **Groupement de services (GDS)** : 1 Frantiq, dans 3 projets.
- **Unité de recherche (UR)** : 12 dans 7 projets (*et UR 4027-Artois dans 2 projets*) et Equipes d'accueil (**EA**) : 17 dans 9 projets (*et EA 4710 -POLEN dans 2 projets*).
- **Maisons des sciences de l'homme (MSH)** : 8 dans 9 projets (*et MISHA dans 2 projets, Dijon dans 2 projets*) + 1 engagement du RnMSH dans 1 projet.
- **Partenaires étrangers** : 23 institutions dans 12 projets.
- **Autres partenaires** : 4 centres d'archives, 9 musées, 1 municipalité, 1 chaire, 1 association, 2 comités, 2 délégations, 1 equipex, 1 fédération, 1 infrastructure de recherche, 1 infra de calcul, 1 programme régional, 5 sociétés savantes.

**Remarque 1** : En 2019, 330 porteurs et partenaires étaient engagés dans les projets déposés, certains d'entre eux plusieurs fois. En 2021, nous comptons 209 engagements comme partenaires, pour 174 entités identifiées.

**Remarque 2** : L'impact de l'appel à projets CollEx-Persée demeure assez fort sur les communautés de l'ESR et du CNRS.

## Actions à réaliser dans le cadre des projets



Ce tableau des actions à réaliser dans le cadre des projets (plusieurs réponses étaient possibles) sont des données extraites des dossiers des déposants. Il s'agit d'une représentation de l'offre de services que l'IS peut développer afin de répondre aux besoins de la recherche.

**Remarque 1 :** les points saillants tournent autour de la numérisation, structuration documentaire et diffusion de corpus et de leur enrichissement. Les actions de valorisation liées aux projets constituent un axe très bien pris en compte par les porteurs.

**Remarque 2 :** trois sujets qui étaient émergents en 2019, restent bien présents : celui lié aux référentiels, alignement et diffusion de métadonnées, le traitement automatique de données et la formation aux outils des humanités numériques.

## Disciplines représentées au sein des projets déposés



D'un point de vue disciplinaire, cet appel à projets est, une fois encore, le reflet du réseau lui-même constitué à hauteur de 25% de délégations documentaires d'excellence en STM, et de 75% de délégations documentaires d'excellence en SHS.

Certaines thématiques se distinguent cette année :

- L'Asie (5 projets)
- L'antiquité et l'archéologie (4 projets)
- Le droit/économie (3 projets)



## Annexe 1 : Lauréats de l'appel à projets CollEx-Persee 2021-2022

Sont indiqués les établissements porteurs administratifs ainsi que les coordinateurs scientifiques des projets (dans l'ordre IST-chercheur). Plus d'informations sur [www.collexpersee.eu](http://www.collexpersee.eu).

1. **BE\_NUM – Bulletin épigraphique numérique** ▪ Université Paris 1 – Panthéon Sorbonne
  - Sébastien DALMON (Université Paris 1) et Jean-Yves STRASSER (UMR 8210 ANHIMA).
2. **CHI-KNOW-PO Corpus – Poésie et économie des savoirs dans la Chine médiévale** ▪ GIP BULAC
  - Soline SUCHET (GIP BULAC) et Marie BIZAIS-LILLIG (Université de Strasbourg).
3. **CHRONOSPEDIA – Création d'une encyclopédie numérique des horloges et des appareils de mesure du temps** ▪ Université Grenoble-Alpes
  - Lucie ALBARET (UGA) et Konstantin PROTASSOV (UGA).
4. **CoBaLPP – Collecte des barèmes législatifs des politiques publiques** ▪ PSE – Ecole d'économie de Paris
  - Colette CADIOU (DipSO – INRAE) et Mahdi BEN JELLOUL (Institut des politiques publiques, PSE – Ecole d'économie de Paris).
5. **CoESciTer – Corpus sur l'enseignement des sciences de la terre. Les chaires de géologie universitaire (19e-20e).** ▪ Université de Lille
  - Geoffrey HARAUX (Université de Lille) et Pierre SAVATON (Université de Caen).
6. **COREL – Code relationnel** ▪ Collège de France
  - Anne CHATELLIER (Collège de France) et Frédéric CONSTANT (Université de Nice Sophia Antipolis).
7. **CorpusSATHMA – Corpus de la sculpture de l'Antiquité tardive et du haut Moyen-Age** ▪ CNRS pour Arar – Archéologie et Archéométrie UMR 5138
  - Blandine NOUVEL (CNRS) et Anne FLAMMIN (CNRS).
8. **DreAM – Dreist ar mor – Dros y môr. Outre-Manche – Accross the Channel. Archives bretonnes, archives galloises : une analyse croisée et spéculaire** ▪ Université de Bretagne Occidentale
  - Philippe LAGADEC (Université de Bretagne Occidentale – CRBC) et Ronan CALVEZ (Université de Bretagne Occidentale – CRBC).
9. **ICB – Itinéraire de la collection Bastaire : des enfants lecteurs aux collectionneurs amateurs** ▪ Université Clermont-Auvergne (pour la MSH)
  - Hélène VEILHAN (Université Clermont-Auvergne, MSH) et Viviane ALARY (Université Clermont-Auvergne, CELIS).
10. **IndexAngkor – Indexation dynamique d'exploration du corpus d'Angkor** ▪ Ecole française d'Extrême-Orient
  - François-Xavier ANDRE (EFEO) et Christophe POTTIER (EFEO).

11. **IRIS – Inventaire Rouge de l'Information Sportive** ▪ Université de Bourgogne
  - *Agnès VIOLA (Université de Bourgogne) et Karen BRETIN-MAFFIULETTI (Université de Bourgogne).*
  
12. **ISSA 2 – Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte (2)** ▪ CIRAD, Centre de coopération internationale en recherche agronomique pour le développement
  - *Anne TOULET (CIRAD) et Franck MICHEL (Université Côte d'Azur, CNRS, Inria).*
  
13. **NumeNord – Plateforme des études nordiques : volet 2.** ▪ Bibliothèque nationale et universitaire
  - *Caroline WOESSNER (Bnu) et Thomas Mohnike (Université de Strasbourg).*
  
14. **NUMISMATES** ▪ Centre d'Etudes Alexandrines (CEAlex)
  - *Marie-Delphine MARTELLIERE (CEAlex) et Thomas FAUCHER (CEAlex).*
  
15. **ORDY GOAL – Oralité Dynamique : Grand Ouest français, Acadie, Louisiane** ▪ Université de Poitiers (pour la MSH)
  - *Michaël NAUGE (Université de Poitiers) et André MAGORD (Université de Poitiers).*
  
16. **ORESM – Œuvres et Référentiels des Etudiants, Suppôts et Maîtres de l'Université de Paris, des écoles et collègues parisiens** ▪ Université Paris 1 – Panthéon Sorbonne
  - *Laurence BOBIS (Université Paris 1) et Thierry KOUAME (Université Paris 1 – LaMOP).*
  
17. **PRET19 – Projet de Répertoire des Emprunteurs et Titres empruntés au XIXe siècle à l'université**
  - Université Paris 1 – Panthéon Sorbonne
    - *Marie-Thérèse PETIOT (Université Paris 1) et Viera REBOLLEDO-DHUIN (Université Paris-Est Créteil).*
  
18. **Rev@ntiq** ▪ CNRS
  - *Marie CHEBANCE (FR 3747 MOM) et Sabine FOURRIER (HiSoMA-UMR 5189).*
  
19. **TariMa – Tarih al-Maghrib. Ecrire l'histoire au Maghreb à l'époque moderne et contemporaine.**
  - GIP BULAC
    - *Benjamin GUICHARD (GIP BULAC) et Antoine PERRIER (CNRS).*

## Annexe 2 : Liste des experts en IST sollicités

Ainsi que le prévoyait l'appel à projets dans le processus d'examen, les dossiers répondant aux critères administratifs d'éligibilité ont fait l'objet d'une pré-évaluation technique par un groupe d'experts en information scientifique et technique, puis d'une évaluation scientifique par le conseil scientifique du GIS CollEx-Persée.

Les personnalités qualifiées intervenant dans l'évaluation et la sélection des dossiers se sont engagées à respecter la confidentialité des projets et à signaler tout conflit d'intérêt éventuel.

Les experts ont été coordonnés par Laurie Aoustet (Université Paris 1), Catherine Désos-Warnier (CollEx-Persée) et Gabrielle Richard (UMS Persée).

1. **ANGEVELLE Catherine**, Université de Lorraine
2. **AYMONIN David**, Agence bibliographique de l'enseignement supérieur
3. **BELLIER Luc**, Université Paris Saclay
4. **BOISSAT Romain**, Maison de l'Orient et de la Méditerranée Jean Pouilloux
5. **BRACCO Laëtitia**, Université de Lorraine
6. **BREUIL Sandrine**, Université de Tours, CESR
7. **CHAUDOREILLES Hélène**, Université Paris 3
8. **DUBOIS Olesea**, Sciences-Po Paris
9. **FORCADELL Sophie**, Sciences Po Paris
10. **GARCIA FERNANDEZ Anne**, Université Grenoble Alpes, ELAN
11. **GEROUDET Marie-Madeleine**, Université de Lille
12. **GUICHARD Benjamin**, GIP BULAC
13. **LEBRE Céline**, Université de Nanterre
14. **LEHOUX Elise**, Université de Paris
15. **LORET Stéphane**, Université de Nantes, ESO
16. **MALLERET Cécile**, Université de Lille
17. **MARLET Olivier**, Université de Tours, UMR CITERES
18. **NAEGELEN Sabine**, Université de Toulouse
19. **OKRET-MANVILLE Christine**, Université Paris Dauphine
20. **OURY Clément**, Muséum national d'histoire naturelle
21. **POUPARDIN Elsa**, GIS URFIST – URFIST de Strasbourg
22. **RAGEOT Laurence**, Université de Tours, MSH Val de Loire
23. **REGE Adeline**, Université de Strasbourg
24. **RIVIÈRE Pauline**, Université Sorbonne Nouvelle Paris 3
25. **SCHWEITZER Jérôme**, Bibliothèque nationale et universitaire
26. **STOLL Mathieu**, MCC-SIAF
27. **STOLL Véronique**, Observatoire de Paris
28. **TARDY Cécile**, Université Paris Nanterre
29. **TISSEAU DES ESCOTAIS Aleth**, Observatoire de Paris
30. **WENZ Romain**, Université de Bordeaux

### BE\_NUM

## Bulletin épigraphique numérique

Université Paris 1 – Panthéon Sorbonne

**Sébastien DALMON (Université Paris 1) et Jean-Yves STRASSER (UMR 8210 ANHIMA).**

Le Bulletin épigraphique (BE), publication essentielle pour la recherche en Antiquité, est une chronique critique sur les inscriptions grecques, de l'époque archaïque jusqu'à l'Antiquité tardive. Devenu un instrument de travail majeur grâce à Jeanne Robert et Louis Robert de 1938 à 1984, il est depuis 1987 rédigé par une équipe internationale de rédactrices et rédacteurs. Le BE paraît annuellement sous la forme d'un article d'environ 200 p. dans la Revue des études grecques (REG). Chaque millésime analyse les publications parues l'année précédente ; ainsi est-il très en avance sur les autres grands instruments de travail annuels en épigraphie, le Supplementum Epigraphicum Graecum et L'Année épigraphique. Cependant, le BE n'est peut-être pas encore aussi largement utilisé qu'il le pourrait. Car ne sont disponibles sur Internet que les PDF de la REG (en accès libre et gratuit sur le portail de Persée et en accès par abonnement sur JSTOR), et les index sont incomplets, non cumulatifs, essentiellement sous la forme imprimée ou de PDF peu maniables. Le projet vise à rendre le BE plus accessible et plus facilement interrogeable, à travers une base de données libre et gratuite qui, à terme, rassemblera toutes les notices depuis 1938 et les index complets. Plusieurs tâches différentes s'imposent : conversion en texte éditable, réalisation de nouveaux index, relectures scientifiques, mise en ligne etc. Il faut pour cela des moyens matériels et humains importants. Le projet déposé porte sur la tranche la plus récente, de 2002 à nos jours. Ultérieurement s'ajouteront la seconde tranche (1987-2001) puis la troisième (1938-1984), qui feront l'objet d'autres demandes de moyens.

English summary : The Bulletin épigraphique is a key-reference for any scholar interested in the Ancient World. It provides a critical analysis of all newly published Greek inscriptions, from the Archaic to the Late Antique Period. It owes its scientific renown to Jeanne Robert and Louis Robert, who wrote it from 1938 to 1984. Since 1987, it has been written by an international team of epigraphists. The BE is published each year as an article of 200 pages or so in the Revue des études grecques. Each issue reviews publications from the year before and is therefore well ahead of other yearly references in epigraphy, such as Supplementum Epigraphicum Graecum and L'Année épigraphique. Nonetheless, the BE is perhaps not yet as widely used as it could be, because it is only available online in the form of PDFs from the issues of REG (through Persée open access and JSTOR access by subscription) and because the indexes are not exhaustive and are available only on paper form or as PFDS. Our project aims at making the BE more easily accessible, as well as searchable, through an open-access and free database gathering all entries from 1938 onwards and including comprehensive indexes. Several tasks will have to be carried out: files have to be converted into editable text, new indexes have to be written, converted texts and indexes have to be checked by professional epigraphists before being put online, etc., for which considerable material as well as human resources are needed. The submitted project only concerns the latest phase of the BE (from 2002). The second (1987-2001) and third (1938-1984) phases will be dealt with subsequently, and will be the subject of further fund requests.

## Poésie et économie des savoirs dans la Chine médiévale

GIP BULAC

**Soline SUCHET (GIP BULAC) et Marie BIZAIS-LILLIG (Université de Strasbourg)**

Le projet CHI-KNOW-PO vise à numériser un corpus d'anthologies poétiques, de commentaires, de dictionnaires et d'encyclopédies de la période médiévale chinoise (ca. 200-1000) et à les traiter par reconnaissance optique des caractères. Il s'agit de constituer un ensemble de données textuelles pour l'étude du rôle de la poésie dans l'économie des savoirs. Celui-ci fait actuellement défaut en format numérique ouvert. Le projet permet la valorisation de fonds chinois anciens conservés en France. Il implique également une démarche d'innovation en entraînant une intelligence artificielle sur un corpus de mises en pages anciennes (organisation verticale d'un texte et de sa glose) afin de fournir des modèles ouverts pour l'apprentissage machine. Ces derniers permettront de faciliter l'acquisition de nouveaux corpus d'imprimés chinois anciens par différents moteurs d'OCR/HTR. Une partie du corpus sera annotée et publiée sur le site du projet CHI-KNOW-PO (en cours de construction) sous forme d'édition enrichie en regard des pages numérisées. Le reste du corpus en plein texte sera mis en ligne, en regard des images correspondantes, dans un but d'enrichissement collaboratif qui se poursuivra au-delà du projet (selon un modèle inspiré par [https:// bonhum.huma-num.fr/](https://bonhum.huma-num.fr/)). L'ensemble du corpus, images et plein texte, sera mis à disposition sous licence ouverte pour des réutilisations libres. Le projet inclut le développement d'un connecteur de NumaHOP vers Nakala. Le projet se conclura par un cycle de conférences réflexives sur les enjeux patrimoniaux et philologiques des humanités numériques.

English summary: The CHI-KNOW-PO project aims at digitizing a corpus made of poetic anthologies, commentaries, dictionaries and encyclopedias from the medieval Chinese period (ca. 200-1000) and at processing them through optical character recognition (OCR). The goal is to offer a set of textual data for the study of the role of poetry in the knowledge economy, corpora that are not widely available yet in an open digital format. The project promotes the scientific use of heritage Chinese collections held in French libraries. The project is innovative, in the way it will train an artificial intelligence on the page layouts of old Chinese printed books, with a vertical organization of the text and its gloss, and will provide open models for machine learning. The latter will facilitate the acquisition of ancient Chinese printed texts by different OCR/HTR engines. A portion of the corpus will be annotated and published on the CHI-KNOW-PO project website (currently under construction) as an enriched edition of the digitized pages. The rest of the corpus in full text will be made available online, with the corresponding pages images, for enhancing collaborative enrichment beyond the project timespan ( following the model of the project <https://bonhum.huma-num.fr/>). The whole corpus, images and full text, will be made available under an open license. The project includes the development of a connector from the management tool for digitization projects NumaHOP to the data repository Nakala. The project will be concluded by conferences on the digital philological issues of digital humanities.

# CHRONOSPEDIA

## Création d'une encyclopédie numérique des horloges et des appareils de mesure du temps

Université Grenoble-Alpes

**Lucie ALBARET (UGA) et Konstantin PROTASSOV (UGA).**

Le projet vise à organiser et diffuser les savoirs sur les horloges, appareils de mesure et d'affichage du temps depuis le 17<sup>e</sup> siècle, et s'inscrit dans l'histoire des méthodes physiques de mesure du temps. L'absence de documents sur l'horlogerie ancienne (pendules, horloges d'édifice, ...) ainsi que celle des formations présentent un risque majeur de perte du savoir-faire horloger qui a fait rayonner la France dans le monde. Pour y remédier, nous proposons de créer une encyclopédie numérique « CHRONOSPEDIA » avec OMEKA-S. Elle sera créée en collaboration avec les grands musées français et européens et réunira les mécanismes principaux des horloges historiques. Les modèles des mécanismes en 3D seront produits par un Maître d'Art horloger de France et inventeur de cette méthode. Le but est de concevoir une bibliothèque numérique collaborative regroupant ces modèles, les archives numériques des musées et des ressources documentaires, celles déjà répertoriées dans les bibliothèques et celles qui ne le sont pas encore, comme les riches archives des associations d'amateurs d'horlogerie. La conception de cette encyclopédie unique au monde sera réalisée avec l'INIST et la BAPSO UGA et la participation des chercheurs dans les méthodes de 3D ou de la réalité virtuelle. Des formations adaptées, à destination de tous les acteurs (conservateurs des musées, personnels des bibliothèques, enseignants-chercheurs, élèves horlogers et membres d'associations partenaires) seront dispensées afin de permettre l'enrichissement collaboratif et la modération des contenus de l'encyclopédie.

English summary : The present project aims to organize and disseminate knowledge about clocks, measuring and time displaying devices since the XVII<sup>e</sup> century and is part of the history of physical methods applied to measure time. The lack of documents on ancient watchmaking (clocks, building clocks, etc.) as well as the lack of training presents a major risk of loss of watchmaking know-how that has made France famous in the world. To remedy this, we propose to create a digital encyclopedia «CHRONOSPEDIA» with OMEKA-S. It will be created in collaboration with major French and European museums and will bring together the main mechanisms of historical clocks. The 3D models of the mechanisms will be produced by a French Master of Watchmaking Art and inventor of this method. The aim of this project is to develop design a collaborative digital library gathering 3D models of mechanisms, digital archives of museums and documentary resources, those already listed in libraries and those not yet listed, like the rich archives of watchmaking amateur associations. The design of this unique encyclopedia will be carried out with the INIST and the BAPSO UGA and the participation of researchers in 3D or virtual reality methods. Adapted training for museum curators, library staff, researchers, watchmaking students and members of partner associations will be provided to enable collaborative enrichment and moderation of the encyclopedia's content.

## Collecte des barèmes législatifs des politiques publiques

PSE – Ecole d'économie de Paris

**Colette CADIOU (DipSO – INRAE) et Mahdi BEN JELLOUL (Institut des politiques publiques, PSE – Ecole d'économie de Paris).**

Ce projet de collecte cherche à répondre à un besoin récurrent et fondamental pour les chercheurs et les administrations publiques de disposer d'une source de référence facilement mobilisable. Il vise ainsi à compléter la base de données des barèmes législatifs des politiques publiques initiée par l'Institut des politiques publiques. D'une part, un effort important de saisie sera entrepris pour améliorer la couverture temporelle des données jusqu'en 1945, tendre vers un corpus plus exhaustif et améliorer les métadonnées, notamment compléter et fiabiliser les références législatives (avec lien vers les textes de lois). D'autre part, des structures permettant la maintenance, la contribution collaborative, et la réutilisation des données seront mis en place, afin de favoriser la diffusion, l'accessibilité et la fiabilité des données. Le développement d'un outil de validation permettra d'établir un diagnostic clair de l'existant pour mieux le compléter. L'introduction d'une interface dédiée permettra d'améliorer la quantité et la qualité des contributions actuelles mais aussi de mobiliser des experts. Afin de faciliter la découvrabilité des données, les séries de paramètres seront indexées et le corpus sera interrogeable à l'aide de son intégration dans un portail DBnomics qui dispose de capacités de visualisation et d'une interface de programmation (API) permettant le téléchargement des séries par les logiciels usuels de statistiques. La base de données comme les outils développés seront sous diffusée sous une licence libre ouverte et seront librement réutilisables.

English summary : This project aims at satisfying a recurring and fundamental need of researchers and policy-makers for a reference source regarding all tax-and-benefit parameters pertaining to French public policy. Its objective is to complete, extend and enhance a database launched by the Institut des politiques publiques. Part of the project consists in an important effort of data collection. This will extend the time coverage of the database back to 1945, aim at a more exhaustive legislative coverage, and strengthen the metadata, by completing legislative references and increase their reliability (links to legislation). The project will also set up tools to simplify database maintenance, collaborative contribution and for thirdparty reusability. This will foster the diffusion, the accessibility, and the overall quality of the database. A validation tool will be developed in order to analyze the existing the database and point towards the most crucial avenues of development. Second, a specific collaborative contribution tool will help increase both the quantity and the quality of third-party contributions from existing partners, and bring in new expert partners. In order to simplify data access and reuse, all parameter series will be indexed and the database will be made accessible via the DBnomics portal. User will benefit from DBnomics visualization and API tools so tax and benefit parameters will be easily downloadable and accessible by common statistical softwares. All products of the project (database and tools) will be released under an open source license and freely reusable.



## CoESciTer

### Corpus sur l'enseignement des sciences de la terre. Les chaires de géologie universitaire (19e-20e)

Université de Lille

**Geoffrey HARAUX (Université de Lille) et Pierre SAVATON (Université de Caen)**

Le projet CoESciTer a pour but d'amorcer la constitution d'un corpus en ligne de sources pour l'histoire de l'enseignement universitaire des sciences de la terre aux 19e et 20e siècles. Le corpus numérique a vocation de proposer des manuscrits de notes de cours prises par des professeurs ou des étudiants ainsi que leur édition électronique en xml-tei, permettant notamment une interrogation du corpus par notion ou objet scientifique. Les financements demandés à CollEx-Persée doivent permettre une montée en compétence au sein du SCD de l'Université de Lille sur la transcription automatisée de manuscrits et l'édition xml-tei. Un corpus initial constitué de cours manuscrits de géologie, paléontologie et minéralogie donnés à l'Université de Lille, avec notamment des figures comme Jules Gosselet ou Charles Barrois, et de cours de paléontologie conservés au Museum d'Histoire naturelle dans le fonds Marcellin Boule, servira à l'expérimentation des méthodologies élaborées dans le cadre du projet. Une enquête menée durant la première phase du projet permettra un état des lieux plus complets des fonds conservés par tous types d'institutions pouvant alimenter le corpus à plus ou moins long terme. Cet état des lieux mentionnera les opérations à prévoir pour une exploitation du fonds (tri, signalement, numérisation...). Une journée d'étude organisée avec le soutien des sociétés savantes s'intéressant à l'histoire des sciences de la terre permettra de présenter le projet à la communauté scientifique, d'en faire le bilan et d'envisager les développements futurs.

English summary : CoESciTer aims to start the constitution of a digitized corpus of research materials about upper education teaching of earth sciences in 19th and 20th centuries. The core of the corpus will be made of manuscripts notes taken by professors or students. Digitized versions will be enriched with transcription and xml-tei edition. Indexes of scientific objects or concepts built with xml-tei beacons will assist researchers in questioning the corpus. CollEx-Persée's funding will allow Lille University Libraries to build up competences in HTR and xml-tei edition. A starting collection will be used to develop a methodology and a workflow. This starting collection will be composed of two sub-corpus : manuscripts courses taught in Lille University by professors like Charles Barrois, Jules Gosselet and others in geology, paleontology and mineralogy between 1870 and 1905 ; manuscripts courses taught by the paleontologist Marcellin Boule in Clermont University and Museum national d'histoire naturelle, with also notes taken by Boule as a student during his time at Toulouse University in the 1880's. Further investigations in the first phase of the project will allow to identify more documents which could join the corpus in middle or long term and operations needed to exploit them. A seminar will be organized to present the project to scientific community with the help of learned societies interested in earth sciences history such as Société géologique de France, Société géologique du Nord and Comité français pour l'histoire de la géologie.



## Code relationnel

Collège de France

**Anne CHATELLIER (Collège de France) et Frédéric CONSTANT (Université de Nice Sophia Antipolis).**

En s'appuyant sur les documents conservés dans les fonds du porteur de la filière IST (Collège de France) et sur le travail accompli dans le cadre de programmes de recherche antérieurs (LSC, EPJ), le présent projet a trois objectifs : 1. La préservation et la valorisation d'un corpus unique de plus de cinq cents titres indispensables pour la recherche sur l'histoire juridique, administrative et sociale de la Chine moderne déposés sur Nakala. 2. La création d'une base de données associant une sélection de textes et d'images de ce corpus. 3. Le développement d'outils de visualisation des relations entre les documents du corpus et les métadonnées qui leur seront associées. Il visera à montrer la faisabilité de l'intégration au sein d'un même flux de travail de différents outils open source. La colonne vertébrale sera constituée par la version texte d'un "code virtuel" retraçant l'évolution du droit de la dynastie des Qing (1644-1911) année après année, qui sera mis en relation avec des jugements appliquant ces lois ou à l'origine de modifications de la législation. Nous extrairons ensuite des sources l'ensemble des métadonnées relatives à ces lois et décisions judiciaires pour constituer une base de données qui permettra, d'une part, d'analyser les conditions de l'évolution du droit chinois de la période impériale tardive, et d'autre part, de construire une ontologie descriptive et relationnelle des sources juridiques et administratives chinoises. Le livrable de cette dernière partie du projet sera un outil de visualisation s'appuyant sur plusieurs variables du droit Qing.

English summary : Based on the documents kept in the collections of the lead institution of the IST program (Collège de France) and on the work accomplished in the course of previous research programs (LSC, EPJ), the present project has three objectives: 1. The preservation and enhancement of a unique corpus of more than five hundred titles essential to research on the legal, administrative and social history of modern China stored on Nakala. 2. The creation of a database associating a selection of texts and images from this corpus. 3. The development of tools for the visualization of the relations between the documents of the corpus and the metadata associated with them. It will aim at showing the feasibility of integrating different open source tools within the same integrated workflow. The backbone of the project will be the text version of a "virtual code", which will trace the evolution of the Qing dynasty's legislation (1644-1911) year after year. It will be linked to judgments based on these laws or which resulted in changes in the legislation. We will then extract from the sources all the metadata related to these laws and judicial decisions in order to set up a database, which will allow us, on the one hand, to analyze the conditions of the evolution of late imperial Chinese law, and on the other hand, to build a descriptive and relational ontology of Chinese legal and administrative sources. The deliverable of this last part of the project will be a visualization tool based on several variables of Qing law.

# Corpus SATHMA

## Corpus de la sculpture de l'Antiquité tardive et du haut Moyen-Age

CNRS pour Arar – Archéologie et Archéométrie UMR 5138

**Blandine NOUVEL (CNRS) et Anne FLAMMIN (CNRS)**

Le projet SATHMA s'est donné pour objectif de reprendre l'inventaire de la sculpture du haut Moyen Âge mené par une équipe du CNRS dans les années 1965-1980 et dont les archives de travail, conservées au musée du Louvre, constituent un fonds patrimonial et scientifique exceptionnel. Seuls quatre volumes du Recueil général des monuments sculptés en France pendant le Haut Moyen Age (IVe–Xe siècles) ont été publiés par le CTHS. Il convient de valoriser ces archives, de même que les données issues d'autres recherches inédites sur ces sculptures (fonds Cabanot...). L'exploitation numérique de cette riche documentation patrimoniale et scientifique, destinée à être versée dans une base de données à interface Wiki consultable en ligne sur la TGIR Huma-Num, offrira un véritable inventaire de sculptures souvent inédites et permettra de la partager à un large public ; elle sera surtout exploitable pour des spécialistes (chercheurs et étudiants). A cette fin, le moteur et l'interface seront mis à niveau des normes d'interopérabilité du web sémantique. Le vocabulaire rassemblé sera aligné sur le thésaurus PACTOLS dont il complètera les branches relatives aux œuvres et au bâti.

English summary : The SATHMA project has set itself the goal of taking over the inventory of early Medieval sculpture carried out by a CNRS team from 1965 until the 1980s and whose working archives, kept at the Louvre Museum, constitute an exceptional fund. Only four volumes of this General Collection of monuments carved in France during the Early Middle Ages (4th–10th cent.) have been published by the CTHS. It is necessary to value these archives, as well as the data from other unpublished researches on these sculptures. (Cabanot funds...). The digital exploitation of this rich heritage and scientific documentation, intended to be uploaded to a Wiki interface database searchable online on the Huma-Num TGIR, will offer a real inventory of often unpublished sculptures and will allow it to be shared to a wide audience; it will be especially usable for specialists (researchers and students). To this end, the engine and the interface will be upgraded to semantic web interoperability standards. The collected vocabulary will be aligned with the PACTOLS thesaurus, thus complementing its branches relating to works and buildings.

## DreAM

# Dreist ar mor – Dros y môr. Outre-Manche – Accross the Channel. Archives bretonnes, archives galloises : une analyse croisée et spéculaire

Université de Bretagne Occidentale

**Philippe LAGADEC (Université de Bretagne Occidentale – CRBC) et Ronan CALVEZ (Université de Bretagne Occidentale – CRBC).**

La bibliothèque nationale du pays de Galles (NLW) possède un important fonds d'archives bretonnes, qui se compose d'archives littéraires, historiques et politiques, et de manuscrits de théâtre en breton. Ces archives trouvent un écho dans des bibliothèques et des centres d'archives français. Il existe donc une complémentarité des fonds, de part et d'autre de la Manche, qui mérite d'être interrogée. Ce projet vise plusieurs objectifs scientifiques. Le premier serait de recenser, numériser, et exploiter les archives bretonnes du pays de Galles, et de les mettre en lien avec les archives qui, en Bretagne ou ailleurs, leur répondent et les complètent : un même travail de valorisation des archives serait donc mené de part et d'autre de la Manche. La numérisation et la production de métadonnées pour les documents sélectionnés permettront d'apporter des réponses à la question de ce qui a prévalu à la constitution et à la situation de ces fonds. C'est d'ailleurs un des rares faits tangibles de relations entre pays dits "celtiques", ce qui justifie que des chercheurs du CRBC et du CAWCS s'y intéressent. Par ailleurs, la réunion de ces fonds et leur accessibilité aux chercheurs permettraient de renouveler la recherche sur certains auteurs, acteurs, périodes ou productions. Défricher ces fonds très souvent inconnus et les exploiter scientifiquement permettra d'affiner grandement l'état des connaissances actuelles sur la littérature bretonne, sur l'histoire des relations littéraires et idéologiques entre Bretagne et pays de Galles, ainsi que sur l'histoire des nationalismes.

English summary: The National Library of Wales (NLW) has a significant Breton archive, consisting of literary, historical, and political archives, as well as manuscripts of historical Breton-language plays. This archive is echoed in French libraries and archives. There is therefore a complementarity of holdings on both sides of the Channel which deserves to be explored. The aims and objectives are as follows. The first would be to identify, digitise and exploit Breton archives in Wales, and to link these with archives in Brittany or elsewhere that respond to and complement them. The same work of enhancing the value of archives would therefore be carried out on both sides of the Channel. Digitisation and the production of metadata for the selected documents will allow for investigation of the underlying reasons for the collections' constitutions and locations. As one of the few tangible facts of relations between so-called "Celtic" countries, this warrants the attention of researchers from the CRBC and the CAWCS. Furthermore, the (digital) reunification of these collections and making them accessible to researchers will open new research avenues in relation to some writers, figures, periods or cultural products. Rigorous scholarly investigation of these often-unknown collections will make a significant contribution to knowledge on Breton literature, on the history of literary and ideological relations between Brittany and Wales, as well as on the history of nationalism.

# Itinéraire de la collection Bastaire : des enfants lecteurs aux collectionneurs amateurs

Université Clermont-Auvergne (pour la MSH)

***Hélène VEILHAN (Université Clermont-Auvergne, MSH) et Viviane ALARY (Université Clermont-Auvergne, CELIS)***

Le Fonds Bastaire est constitué de deux ensembles documentaires distinctifs. La collection est riche d'environ 10.000 livres, journaux et titres de revues illustrées publiés en France entre 1850 et 1940, à l'apogée du roman populaire. Elle a été collectée au tournant des années 1960 par Jean et Michel Bastaire qui ont entrepris de reconstituer la bibliothèque de leur enfance. Les archives regroupent notamment les archives d'enfance de Jean Bastaire ainsi que ses archives de collectionneur et ses archives personnelles d'écrivain et de publiciste. Le projet a pour but de rendre visible l'articulation entre la collection et les archives en faisant du catalogue de la collection le point d'entrée dans le corpus. L'objectif documentaire est de numériser et rendre ce matériel disponible pour la recherche. Le projet scientifique consiste à étudier le processus de constitution d'une collection en montrant le cheminement de l'enfant lecteur au collectionneur amateur et à l'écrivain spécialisé. Ce parcours sera exploré autour de trois axes qui permettront d'interroger le concept de culture populaire tel qu'il s'exprime à la fois dans la relation qui se noue entre le texte et l'image, dans un schéma sociogéographique familial exemplaire et dans le modèle du collectionneur.

English summary : "Fonds Bastaire" is a corpus of printed texts and archives made up of two separate sets of documents. The collection of printed texts contains around 10.000 books, newspapers and illustrated magazine titles published in France between 1850 and 1940, in the heyday of the popular novel. It was amassed at the turn of the 1960s by Jean and Michel Bastaire, who set out to reassemble the library of their childhood. The collection includes in particular Jean Bastaire's childhood and collector archives, as well as his personal archives as a writer and journalist. The aim of the project is to uncover the link between the collection and the archives by using the catalog of the collection to give access to the corpus. From a documentary point of view, the objective is to digitize this material and make it available for research. The scientific project consists in studying the process of assembling a collection by showing the progress from the young reader to the amateur collector and the specialized writer. This evolution will be explored around three axes that will allow us to analyze the concept of popular culture as it is expressed in the relationship formed between text and image, in an exemplary social and geographic family context, and in the image of the collector.

# IndexAngkor

## Indexation dynamique d'exploration du corpus d'Angkor

Ecole française d'Extrême-Orient

***François-Xavier ANDRE (EFEO) et Christophe POTTIER (EFEO)***

Les archives de la Conservation d'Angkor (1907-1975) sont un patrimoine documentaire unique dont la publication numérique complète est très attendue par les communautés de chercheurs, au Cambodge comme au niveau international, mais aussi par le grand public. L'ensemble du fonds, constitué des Journaux de fouilles, des rapports du Conservateur, des photographies et des fiches d'objets, est conservé à l'EFEO. Le projet IndexAngkor se propose d'achever la numérisation des secteurs documentaires restants à numériser (photos de la conservation et certaines catégories de fiches du dépôt). En parallèle, il propose le développement d'outils Open source de navigation au sein du corpus pour permettre à l'utilisateur de passer sagement d'un type de document à l'autre. Ces fonctionnalités seront basées sur le standard IIIF et implémentées sous la forme d'un module Omeka S. Elles permettront la constitution d'index dynamiques des Journaux de fouilles de la Conservation.

English summary: The archives of the Angkor archaeological department (1907-1975) are a unique documentary heritage whose complete digital publication is eagerly awaited by the research community, both in Cambodia and internationally, as well as by the general public. The entire collection, consisting of excavation journals, curatorial reports, photographs, and object files, is held at the EFEO. The IndexAngkor project proposes to complete the digitization of the remaining documents to be digitized (photos and certain categories of objects records). In parallel, it proposes the development of open source navigation tools within the corpus to allow the user to move smoothly from one type of document to another. These functionalities will be based on the IIIF standard and implemented in the form of an Omeka S module. They will allow the constitution of dynamic indexes of the Conservation's excavation logs.

## Inventaire Rouge de l'Information Sportive

Université de Bourgogne

***Agnès VIOLA (Université de Bourgogne) et Karen BRETIN-MAFFIULETTI (Université de Bourgogne).***

Mobilisant l'expertise de la MSH de Dijon en matière de conservation et diffusion de corpus documentaires, le projet IRIS (Inventaire « Rouge » de l'Information Sportive) s'attachera à traiter et exploiter la collection complète du mensuel Miroir du cyclisme (1960-1994). Une équipe pluridisciplinaire observera, par une étude de contenus (titres, articles, iconographie, photographies), le discours médiatique (et critique) porté sur le sport par un organe de presse proche du Parti communiste français. Et ce à partir d'entrées classiques : les compétitions internationales et championnats nationaux, courses et critères régionaux, figures de champions, formes de présentation du spectacle sportif, etc. Au-delà, la mobilisation de méthodologies innovantes et adaptées au traitement de séries longues permettra de porter une attention toute particulière aux formes de l'écriture journalistique et à l'éventuelle identification d'un patrimoine linguistique commun qui viendrait s'agréger à l'ensemble des sphères de la « galaxie communiste », par-delà la seule presse. L'analyse lexicale et syntaxique, le traitement iconographique, la fouille de données, le comparatisme sont les principales méthodes retenues, en sus des formes traditionnelles de traitement des sources par l'historien. S'inscrivant dans une historiographie pionnière s'agissant de l'histoire de la presse sportive, ce projet offre l'occasion d'une réflexion ciblée sur les usages politiques d'un sport populaire emblématique, pour le temps contemporain.

English summary : Mobilizing the expertise of the MSH of Dijon in the field of conservation and dissemination of documentary corpus, the IRIS project (Inventaire " Rouge " de l'Information Sportive) will treat and exploit the complete collection of the monthly magazine Miroir du cyclisme (1960-1994). A multidisciplinary team will observe, through a study of content (titles, articles, iconography, photographs), the media discourse (and criticism) on sport, by a press organ close to the French Communist Party. Usual focuses for reflection will be adopted : international competitions and national championships, races and regional criteriums, champions' figures, forms of presenting the sports show, etc. Beyond this first approach, the mobilization of innovative methodologies adapted to the treatment of long series will allow to pay particular attention to the forms of journalistic writing and to the possible identification of a common linguistic heritage. And if so, this one could join other fields of the "communist galaxy". Lexical and syntactic analysis, iconographic processing, data mining, and comparative history are the main methods used, in addition to the traditional forms of source processing by the historians. As part of a pioneering historiography on the history of the sports press, this project offers the opportunity of focused reflections on the political uses of a popular sport emblematic of contemporary times.

## ISSA2

### Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte (2)

CIRAD, Centre de coopération internationale en recherche agronomique pour le développement

*Anne TOULET (CIRAD) et Franck MICHEL (Université Côte d'Azur, CNRS, Inria).*

S'inscrivant dans la dynamique de la science ouverte et le respect des principes FAIR, le projet ISSA 2 propose une méthode générique pour explorer les données disponibles dans une archive ouverte afin 1) d'en extraire des connaissances nouvelles, 2) d'exploiter ces connaissances avec un objectif bibliométrique et 3) de proposer des services aux chercheurs et aux documentalistes, en particulier en termes de bibliométrie et de recherche d'information. Le projet s'appuie sur l'index sémantique produit dans le cadre du projet ISSA, décrivant les publications d'une archive ouverte : métadonnées, descripteurs et entités nommées mentionnées dans le texte, liés à des bases de connaissances standard. Les méthodes proposées exploitent des techniques de fouille de données ainsi que des techniques de construction, publication et exploitation de graphes de connaissances (Web sémantique et Web des données). L'apport de l'intelligence artificielle doit permettre d'enrichir et élargir les possibilités offertes par des indicateurs bibliométriques par ailleurs bien connus. Deux archives ouvertes serviront de cas d'utilisation tout au long du projet : Agritrop, l'archive ouverte du Cirad, et l'instance HAL de l'UR EuroMov Digital Health in Motion.

English summary : In line with open science and the FAIR principles, the ISSA 2 project proposes a generic method to explore the data available in an open archive in order to 1) extract new knowledge, 2) exploit this knowledge with a bibliometric objective and 3) propose services to researchers and documentalists, in particular in terms of bibliometrics and information retrieval. The project is based on the semantic index produced in the context of the ISSA project, describing the publications of an open archive: metadata, descriptors and named entities mentioned in the text, linked to standard knowledge bases. The proposed methods exploit data mining techniques as well as techniques for the construction, publication and exploitation of knowledge graphs (Semantic Web and Web of Data). The contribution of artificial intelligence should allow us to enrich and expand the possibilities offered by bibliometric indicators which are otherwise well known. Two open archives will serve as use cases throughout the project: Agritrop, the CIRAD open archive, and the HAL instance of the UR EuroMov Digital Health in Motion.

## Plateforme des études nordiques : volet 2

Bibliothèque nationale et universitaire

***Caroline WOESSNER (Bnu) et Thomas Mohnike (Université de Strasbourg).***

Ce projet représente une évolution de l'offre de service actuelle de la plateforme des études nordiques déjà en ligne grâce à un financement CollEx-Persée obtenu en 2018. Il s'agit de mieux répondre aux attentes des chercheurs en développant une nouvelle dimension de services autour de de l'exploitation de corpus numériques et de la valorisation des résultats de la recherche, au moyen d'outils et services en lien avec les humanités numériques et les données bibliométriques sous la forme d'un environnement de travail intégré. Il s'agit aussi de renforcer la dimension internationale du réseau constitué, en s'appuyant sur notre collaboration avec la plateforme allemande ViFaNord et ses partenaires parmi les bibliothèques scandinaves. A terme, c'est une offre de service à dimension européenne qui pourra émerger des initiatives lancées autour de cette version augmentée de la plateforme des études nordiques.

English summary : This project represents an evolution of the current service offer of the Nordic Studies platform already online thanks to a CollEx-Persée funding obtained in 2018. The aim is to better meet the expectations of researchers by developing a new dimension of services around the exploitation of digital corpora and the valorisation of research results, by means of tools and services linked to digital humanities and bibliometric data in the form of an integrated work environment. The aim is also to strengthen the international dimension of the network, based on our collaboration with the German platform ViFaNord and its partners among Scandinavian libraries. In the long run, a service offer with a European dimension could emerge from the initiatives launched around this enhanced version of the Nordic Studies platform.



# NUMISMATES

Centre d'Etudes Alexandrines (CEAlex)

**Marie-Delphine MARTELLIERE (CEAlex) et Thomas FAUCHER (CEAlex).**

L'Institut français d'archéologie orientale (Ifao) conserve, dans ses archives, une base de données manuscrite unique, référencant l'ensemble des numismates, établie au Caire par Marcel Jungfleisch dans les années 1930-40. Ces « notices nécrologiques » comportent plus de 17000 fiches pour plus de 16000 individus couvrant le monde entier à partir du 13e siècle jusqu'au début du 20e siècle. Profitant d'un premier travail financé par l'Ifao, le projet a pour but de mettre en ligne cette base de données en bénéficiant à la fois des projets existants en matière d'archives numismatiques (<https://fina.oeaw.ac.at/#/de/start>) et de l'ontologie déjà développée pour la discipline (<http://nomisma.org>). Outre les informations sur la vie de chaque numismate, la base de données comprendra plusieurs modules dont deux permettront d'avoir accès d'un côté aux archives (épistolaires, administratives, etc) de chaque individu et de l'autre à sa collection de monnaies (vente, don, etc), en bénéficiant dans un premier temps de l'apport des informations présentes au département des Monnaies, médailles et antiques de la BnF ainsi qu'aux archives du CEAlex. Chaque entrée sera liée à d'autres bases de données référentielles (IdRef, VIAF, ISNI, Wikidata, Calames) fournissant ainsi un outil indispensable aux historiens et aux collectionneurs pour reconstruire l'histoire de la numismatique et de l'antiquarianisme.

English summary : The Institut français d'archéologie orientale possesses, in its archives, a unique handwritten database on all the numismatists that was made in Cairo by Marcel Jungfleisch during the years 1930-40. This series of obituaries consist of more than 17000 paper cards for more than 16000 individuals from all over the world since the 13th century until the beginning of 20th century. Benefiting from a first work founded by the Ifao, the project is aiming at putting online this database taking into advantage the existing projects in the field, may it be from epistolary archives (<https://fina.oeaw.ac.at/#/de/start>) or from the ontology already developed for numismatics (<http://nomisma.org>). Besides information on the life of each numismatist, the database will contain different modules of which two will give access on one side to archives (epistolary, administrative, etc) of each individual and to coin collections (sale, donation, etc), benefiting first from the data coming from the Department of Coins of the French National Library as well as from the archives of the CEAlex. Each entry will be linked to other referential databases (IdRef, VIAF, ISNI, Wikidata, Calames) providing therefore a powerful tool for Historians and to Collectors to reconstruct the History of Numismatics and Antiquarianism.

# ORDY GOAL

## Oralité Dynamique : Grand Ouest français, Acadie, Louisiane

Université de Poitiers (pour la MSH)

**Michaël NAUGE (Université de Poitiers) et André MAGORD (Université de Poitiers).**

Oralité Dynamique : Grand Ouest français, Acadie, Louisiane Grâce au soutien CollEx-Persée au projet Acadie Sonore (lauréat de l'AAP 2019-2020), nous avons pu expérimenter et monter en compétence sur la « FAIRisation » des archives sonores du Grand-Ouest français et d'Amérique du Nord francophone. Ce travail autour de corpus mis en commun a permis d'initier les questions scientifiques avec les chercheurs internationaux du projet autour la spécificité de l'oralité dans l'étude du lien entre le culturel et le social. Fort de l'expérience acquise et des réussites obtenues, nous envisageons dans le cadre du projet « Oralité Dynamique » un développement structurant de l'internationalisation des actions techniques et scientifiques. Cette évolution de la dimension collaborative à l'échelle de deux continents s'appuiera sur l'organisation d'une gouvernance fluide afin de fédérer chaque équipe de techniciens et chercheurs, sur la montée en puissance de dépôts par lots (amélioration continue de l'enrichissement des collections et des fonctionnalités de la bibliothèque sonore grâce aux transferts réciproques de compétences), sur la mise en œuvre d'un programme scientifique novateur et enfin sur le développement d'un indispensable programme de valorisation et de médiation très ambitieux. Afin d'effectuer une diffusion consistante des données auprès de publics diversifiés, deux nouveaux partenariats seront actés avec les deux structures de culture et de médiation scientifiques phares en Nouvelle Aquitaine. Le projet « Oralité Dynamique », unique dans les collections qu'il réunit, dans ses objectifs scientifiques et dans son programme de restitution-valorisation auprès de publics cibles, pourra ainsi pleinement lancer sa dynamique collaborative internationale, qui ne manquera pas, par effet d'entraînement, d'élargir encore le réseau.

English summary : Oral dynamics : Western France, Acadia, Louisiana Thanks to CollEx-Persée's support for the Acadie Sonore project (winner of the 2019-2020 AAP), we were able to experiment and gain expertise on the "FAIRization" of sound archives in the French Grand-Ouest region and in Frenchspeaking North America. This work around a pooled corpus made it possible to initiate scientific questions with the international researchers of the project around the specificity of orality in the study of the link between the cultural and the social. Building on the experience acquired and the successes obtained, we are considering, within the framework of the "Dynamic Orality" project, a structuring development of the internationalization of technical and scientific actions. This evolution of the collaborative dimension on the scale of two continents will be based on the organization of a fluid governance in order to federate each team of technicians and researchers, on the rise in importance of batch repositories (continuous improvement of the enrichment of the collections and functions of the sound library thanks to the reciprocal transfers of skills), on the implementation of an innovative scientific program and finally on the development of an essential program of promotion and very ambitious mediation. In order to disseminate the data consistently to diverse audiences, two new partnerships will be entered into with the two leading scientific culture and mediation structures in New Aquitaine. The "Dynamic Orality" project, unique in the collections that it brings together, in its scientific objectives and in its restitution-promotion program for target audiences, will thus be able to fully launch its international collaborative dynamic, which will not fail, by effect of training, to further expand the network.

## Œuvres et Référentiels des Etudiants, Suppôts et Maîtres de l'Université de Paris, des écoles et collègues parisiens

Université Paris 1 – Panthéon Sorbonne

*Laurence BOBIS (Université Paris 1) et Thierry KOUAME (Université Paris 1 – LaMOP)*

Le projet de portail web agrégatif ORESM vise à accompagner les recherches sur l'Université de Paris, les écoles et les collèges parisiens au Moyen Âge (1200-1600). Il proposera des référentiels personnes physiques et morales du Moyen Âge, élaborés à partir de l'analyse des données de la base Studium Parisiense, des données issues de dépouillements réalisés dans le cadre du projet et des référentiels des Archives nationales. Une application web permettra d'aligner les données de différentes bases prosopographiques, de réconcilier les identités des personnes et de les relier à des référentiels externes (ABES, Biblissima, IRHT), tout en gérant l'incertitude. Un inventaire dynamique virtuel des archives actuellement dispersées entre plusieurs institutions se matérialisera dans une base de données en graphe, alimentée par les descriptions des pièces d'archives et reliant les entités entre elles (personnes, archives, lieux, etc.) via l'application web. Le graphe sera interrogeable grâce aux technologies sémantiques (SPARQL). Une interface de recherche et de consultation des données offrira des outils d'accès aux référentiels. Elle permettra l'exploration et la consultation dynamique de l'inventaire des archives et renverra à leurs descriptions. De plus, des APIs ouvriront un accès machine distant. Le graphe et son interface de recherche seront produits grâce aux technologies du web sémantique (RDF, RDFS, OWL). La principale ontologie de référence sera Records in Contexts-Ontology (RiC-O). Les données seront conformes aux principes FAIR.

English summary : The ORESM aggregative web portal project aims to support research on the University of Paris, the Parisian schools and colleges in the Middle Ages (1200-1600). It will provide repositories for individuals and corporate bodies from the Middle Ages, based on the analysis of data from the Studium Parisiense database, data from the research carried out as part of the project and the repositories of the National Archives. A web application will make it possible to match data from different prosopographical databases, to reconcile the identities of individuals and to link them to external reference systems (ABES, Biblissima, IRHT), while managing uncertainty. A virtual dynamic inventory of the archives currently dispersed among several institutions will materialise in a graph database, fed by the descriptions of the archival items and linking the entities together (persons, archives, places, etc.) via the web application. The graph will be searchable using semantic technologies (SPARQL). A data search and consultation interface will provide tools for accessing the repositories. It will allow the exploration and dynamic consultation of the archive inventory and will refer to their descriptions. In addition, APIs will provide remote machine access. The graph and its search interface will be produced using semantic web technologies (RDF, RDFS, OWL). The main reference ontology will be Records in Contexts-Ontology (RiC-O). The data will be in accordance with FAIR principles.

## PRET19

# Projet de Répertoire des Emprunteurs et Titres empruntés au XIXe siècle à l'université

Université Paris 1 – Panthéon Sorbonne

**Marie-Thérèse PETIOT (Université Paris 1) et Viera REBOLLEDO-DHUIN (Université Paris-Est Créteil)**

PRET19 est un projet de numérisation et valorisation scientifique des registres de prêt des bibliothèques de la Sorbonne, de l'École normale supérieure et de la Bibliothèque Sainte-Geneviève au XIXe siècle. Centré sur le public universitaire parisien et son activité d'emprunteur dans ces trois bibliothèques du Quartier latin, le projet a pour objectif de faciliter l'exploration de ces sources manuscrites d'accès relativement confidentiel, en vue d'alimenter la recherche sur l'histoire des bibliothèques universitaires, leur rôle dans la circulation du livre d'érudition et de recherche, ainsi que l'étude du public savant et de ses pratiques d'emprunt dans le contexte de transformation du travail universitaire et d'institutionnalisation des disciplines académiques, au cours du XIXe siècle. Ce projet entend favoriser l'accès des chercheurs aux registres des trois bibliothèques via un site unique qui permettra tout d'abord d'en consulter les versions numérisées et d'en interroger le contenu par date, par nom d'emprunteur et par différents critères résultant d'un enrichissement de la description des emprunteurs. Le projet prévoit en outre l'identification et l'enrichissement des données bibliographiques permettant une interrogation par document emprunté. Ce travail, plus complexe sera réalisé en différentes phases. La première phase, incluse dans ce projet de 24 mois, portera prioritairement sur les échantillons et enquêtes ciblées définies par les partenaires du projet. Le projet est conçu pour s'ouvrir à d'autres bibliothèques partageant le même type de sources et de public.

English summary: The PRET19 project focuses on the 19th century Parisian academic audience and its borrowing activity in three libraries of the Latin Quarter: the Sorbonne, the École normale supérieure and the Sainte-Geneviève Library. Digitization and scientific valorization of the loan registers aim at facilitating the exploration of those relatively confidential manuscript sources, providing materials for research into the history of university libraries, their role in the circulation of scholarly books, and the borrowing practices of the academic audience, in the context of the transformation of academic work and disciplines during that time period. Its goal is to provide researchers with access to the digitized registers of the three libraries via a single site, making it possible to consult them and to search their contents by date, by borrower's name, and by various criteria resulting from the data enhancement regarding the borrowers' descriptions. It will also include the identification and enrichment of bibliographic data, allowing for a search by borrowed document. This more complex part of the work will be carried out in different phases. The first phase, covered in this 24-month project, will focus on samples and targeted surveys defined by members of the project's scientific committee. The project is designed to be open to other libraries sharing the same type of sources and audiences, for interoperability and partnerships.

CNRS

**Marie CHEBANCE (FR 3747 MOM) et Sabine FOURRIER (HiSoMA-UMR 5189)**

L'accès à l'intégralité des collections de revues en SHS, dispersées sur de nombreux portails, est crucial pour la recherche. La connaissance des indicateurs et des politiques éditoriales des revues est également nécessaire pour orienter les stratégies de publication des chercheurs et les engager vers la science ouverte. Aucun outil de ce type n'existe pour les revues en archéologie et sciences de l'Antiquité. Rev@ntiq a l'ambition de construire un service complet de connaissance sur les revues du domaine, qui offre à la fois un accès unique et facilité aux collections numériques et une aide au choix des stratégies de publication. Dans ce but, le projet fournira les moyens humains et techniques pour enrichir et faire évoluer la plateforme Mir@bel, qui centralise l'accès aux contenus en ligne de périodiques, et pour la lier à d'autres bases de connaissance. Un processus d'évaluation des besoins et de phases de tests auprès d'un panel de chercheurs permettra de développer et d'adapter le service. La méthode d'évaluation des besoins et des processus de choix et d'intégration des données sera décrite dans un guide méthodologique à des fins de répliquabilité. Grâce au partenariat avec le réseau Frantiq et les acteurs du Plan de Conservation Partagée des périodiques en sciences de l'Antiquité et archéologie, les professionnels de l'IST s'impliqueront durablement dans le suivi pérenne de ce service d'aide à la consultation et au choix du support de publication. Mir@bel qui expose ses données en accès ouvert permettra également de valoriser et partager largement ces nouvelles métadonnées.

English summary : Access to the entire collections of journals in SSH, disseminated on many portals, is crucial for research. Knowledge of journal indicators and editorial policies is also necessary for publication strategies in order to comply with open science guidelines. However, for journals in archaeology and sciences of antiquity, no tool brings together all these data. Rev@ntiq aims to build a hybrid knowledge service on these journals by combining easier access to digital collections and assistance in the choice of publication strategies. For this purpose, the project will provide the human and technical resources needed to enrich and develop the current Mir@bel platform, which centralizes access to online content of journals, and to link it to other knowledge bases. A needs assessment and user tests with a panel of researchers will enable the service to be adapted during its development, if necessary. The methodology followed for needs assessment, prioritization of periodicals, data collection and integration into the platform will be described in a methodological guide to ensure replicability. Thanks to the partnership with the Frantiq network and the actors of the Plan de Conservation Partagée des périodiques in sciences of Antiquity and archaeology, librarians will be involved in the long-term monitoring of the developed service. Mir@bel website, which makes all its content freely available and reusable, will contribute to the promotion and dissemination of the collected information on the journals.

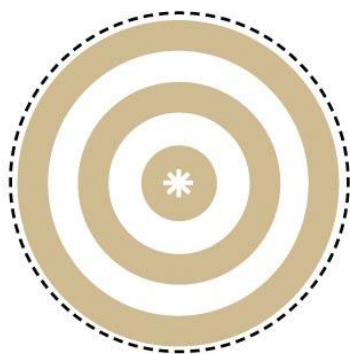
## Tarih al-Maghrib. Ecrire l'histoire au Maghreb à l'époque moderne et contemporaine

GIP BULAC

**Benjamin GUICHARD (GIP BULAC) et Antoine PERRIER (CNRS)**

Le projet Tarima associe le développement des techniques de reconnaissance des caractères de manuscrits (HTR) et imprimés (OCR) de l'arabe maghrébin et l'exploitation par la recherche des collections d'ouvrages maghrébins conservés dans des bibliothèques françaises. Son but est de mettre à disposition la numérisation au format image et au format texte de ces documents jusqu'à présent peu ou pas exploités. Leur mise en ligne contribuera à la valorisation des sources arabes, au cœur des renouvellements récents de l'histoire du Maghreb. En fédérant une communauté de professionnels, de chercheurs et d'étudiants, il proposera un modèle de numérisation enrichie et de visualisation enrichie de l'arabe reproductible sur d'autres corpus. Le corpus sélectionné est composé de manuscrits mais aussi de textes imprimés relevant du genre historique (chronique, récits de voyage, dictionnaires biographiques, histoire universelle, généalogie, autobiographie...) dans un sens très large, genre négligé dans l'étude des manuscrits arabes au profit des sciences religieuses ou de la linguistique. Après la numérisation et l'extraction de texte, un travail sur l'encodage et la valorisation des manuscrits sera conduit avec des chercheurs et des étudiants au sein de la Bulac. Ce même public sera invité à participer à des stages à Rabat et à Tunis de découverte des collections locales de manuscrits, le projet s'achevant par un colloque international sur les usages numériques des sources arabes. Le projet contribuera ainsi à développer les échanges scientifiques internationaux autour des sources textuelles maghrébines.

English summary : The Tarima project combines the development of character recognition techniques for manuscripts (HTR) and prints (OCR) of Maghrebi Arabic texts and the scientific study of collections of Maghrebi works held in French libraries. It aims to make available the digitization in image and text format of a set of primary sources, that have been little used until now. Making them available online will promote the use of Arabic sources, which are at the heart of recent shifts in the historiography of the Maghreb. By federating a community of professionals, researchers and students, it will propose a model of enriched digitization and enriched visualization for Arabic texts reproducible on other corpora. The selected corpus is composed of manuscripts but also of printed texts among the historical genre (chronicles, travelogues, biographical dictionaries, universal history, genealogy, autobiography...) in a very broad sense. Historians using Arabic manuscripts had neglected the study of this genre, in comparison with religious sciences or linguistics. After the digitization and extraction of the text, researchers and students will work on the encoding and enhancement of the manuscripts at the Bulac. The same public will participate in courses in Rabat and Tunis to discover the local collection of manuscripts. The project will end with an international colloquium on the digital uses of Arabic sources. The research project will contribute to develop international scientific networks for the study of maghrebi texts.



## En savoir plus sur l'AAP 2021-2022 :

[collexpersee.eu](http://collexpersee.eu) , rubrique « **Projets** ».

Pour suivre notre actualité,  
rendez-vous sur Twitter  
**@COLLEX\_IR**