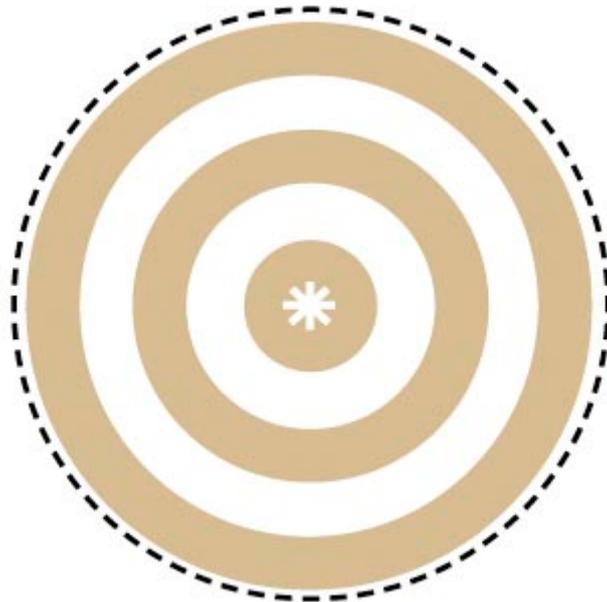


Rapport d'évaluation

Conseil scientifique du GIS
du 21/09/2020



Au terme d'un appel à projets ouvert le 1^{er} octobre 2019 et clos le 15 janvier 2020, 46 dossiers ont été déposés auprès du GIS CollEx-Persée. Une évaluation en plusieurs étapes a été organisée. A l'issue du processus, le conseil scientifique du GIS a désigné, en séance plénière du 18 mai 2020, les lauréats. Les résultats ont été présentés au conseil des membres du GIS, le 20 mai 2020. Ils sont à retrouver sur le web : https://www.collexpersee.eu/aap19_20-resultats/.

25 projets ont été financés à hauteur de 1.997.886 euros.

Le conseil scientifique souligne le haut niveau des dossiers déposés dans ce deuxième appel à projets CollEx-Persée, et considère que ce mécanisme, destiné à favoriser les changements de pratiques sur le terrain, remplit son objectif. Caractéristiques des projets retenus :

- **Co-construction réelle des projets**, entre professionnels de l'IST d'une part, et équipes de recherche, d'autre part. Il est à noter que CollEx-Persée est le seul dispositif, aujourd'hui, à favoriser ce type de portage. Les modalités de collaboration entre les partenaires issus de cultures différentes sont équilibrées et les objectifs des propositions sont bien définis, communs et partagés.
- **Solidité et profondeur scientifique du corpus** sur lequel s'appuie le développement du projet. Les projets exploitent une typologie de matériaux documentaires de natures diverses : livres d'artistes, images, archives sonores, données, référentiels, périodiques, thèses, etc. Tous présentent des défis de traitement et d'exploitation importants, qui, une fois relevés, seront utiles à une large communauté.
- **Diversité des typologies de traitement et de valorisation** dans le cycle de production et d'exploitation du corpus proposé, ce qui répond à ce que souhaitait précisément encourager cet appel. Sont ainsi proposés la conception d'outils d'analyse de données massives, l'océrisation de données chiffrées, l'automatisation de l'océrisation de sources manuscrites, la création de modèles d'apprentissage pour l'OCR, la mise en place de plateformes de recherche collaborative, le dispositif d'accompagnement des chercheurs en matière de gestion de données, l'indexation sémantique des publications dans une archive ouverte, l'expérimentation de nouvelles méthodes de visualisation, etc.
- **Prise en compte des recommandations de la science ouverte**, notamment dans la diffusion des livrables et la garantie d'accès et de pérennité pour la communauté scientifique. Cela suppose le respect des normes, standards et référentiels internationaux dans la gestion des données concernées. Mais aussi la bonne prise en compte des **aspects juridiques** liés au corpus et à sa diffusion. Dans ce cadre, CollEx-Persée ne se substitue pas à d'autres organismes œuvrant dans ce domaine (CoSO, Huma-Num, CCSD/HAL) mais agit en complémentarité de leurs offres et outils, à partir des collections d'excellence pour la recherche.
- **Complémentarité et expérience des équipes impliquées au-delà de la seule institution porteuse**. Certains projets qui n'étaient pas assez mis en réseau ont été écartés. Par ailleurs, si le conseil scientifique note qu'il y a davantage d'institutions étrangères parties prenantes dans les projets qu'en 2018, cela mérite d'être encore encouragé. Certains projets s'y prêtent et pourraient se positionner par rapport à des initiatives internationales.

- **Projets présentant des caractères avérés d'innovation, d'expérimentation, d'exploration, susceptibles de servir une communauté scientifique élargie.** La méthodologie mise en place et les outils utilisés, voire créés, ont le plus souvent la caractéristique d'être transposables à d'autres projets. Le conseil scientifique a été particulièrement attentif à éviter un effet de dispersion des financements sur des projets trop institutionnels, au bénéfice d'une stratégie scientifique susceptible de profiter au plus grand nombre.
- **Impact** sur des publics bien identifiés, au-delà même d'un premier cercle de « confort ». La plupart des projets ont aussi proposé un volet étoffé de **valorisation, formation et communication** en vue de partager et diffuser les résultats. Certains d'entre eux comportaient, en outre, un volet science participative.
- **Savoir-faire en ingénierie de projet** comportant une gouvernance, un calendrier réaliste et un budget cohérent. Parmi les projets non retenus, un fort besoin d'accompagnement en ingénierie de projet a été noté. C'est un point auquel le dispositif est sensible et qu'il pourrait décider de soutenir selon des modalités à mettre en place avec certains de ses partenaires, notamment des organismes de formation.

Cette expérience, importante pour l'ensemble de la communauté scientifique et de l'information scientifique et technique, mériterait d'être prolongée par un troisième appel à projets d'ici la fin du GIS, en 2022.

Ce mécanisme est salué par la communauté de chercheurs comme complémentaire d'autres dispositifs existants et assez unique en son genre. Il a agi comme un levier de transformation des pratiques IST inaugurant un véritable rapprochement entre chercheurs, bibliothécaires, opérateurs, faisant de ces deux derniers de véritables acteurs des projets de recherche. Son impact est non négligeable car plus de 330 porteurs et partenaires se sont engagés dans ces projets.

Cependant, il est nécessaire, au préalable, de réunir cette communauté, en vue de capitaliser sur les résultats afin de déterminer des axes communs susceptibles de revêtir une dimension nationale soutenus par le dispositif.

CollEx-Persée est désormais ce lieu susceptible de structurer et d'accompagner des communautés d'innovation, en associant les opérateurs nationaux. Suite aux deux appels à projets 2018 et 2029, CollEx-Persée soutient et accompagne 6 boursiers et 59 projets pour un montant global de 3.586.557 euros.

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

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

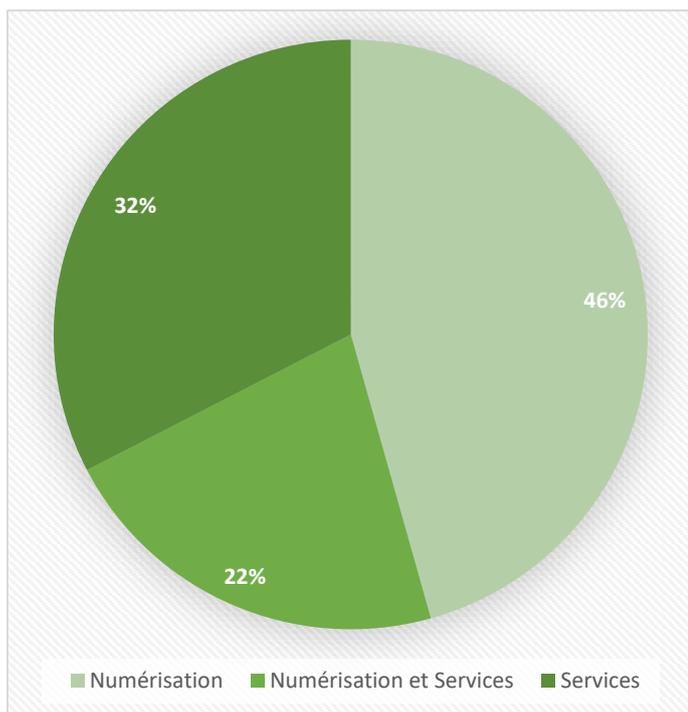
➔ Les attendus et critères de l'appel à projets sont à retrouver :

<https://www.collexpersee.eu/appel-a-projets/>

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 lauréats.....	8
Annexe 1 : Liste des lauréats	9
Annexe 2 : Liste des experts en IST sollicités.....	11
Annexe 3 : Résumés en français et en anglais des projets retenus.....	12
ABRICO 2	12
Acadie Sonore	13
AMOr	14
ArchéoRef Alignements	15
ArchivesBaelo	16
Archivos 3.0	17
BaOIA	18
BibracteVilleOuverte	19
Chronika	20
CLA++	21
CLIODIFLE	22
CONVEX 2	23
Datacc'+	24
Datapoc 2.0	25
DISSIVENT	26
ÈS LETTRES	27
Familielettres	28
FemEnRev	29
ISSA	30
La guerre du droit	31
METADICTMEDICA	32
NumBois	33
PALiBr	34
POPP	35
Transcrire	36

Statistiques de dépôts et typologie des projets

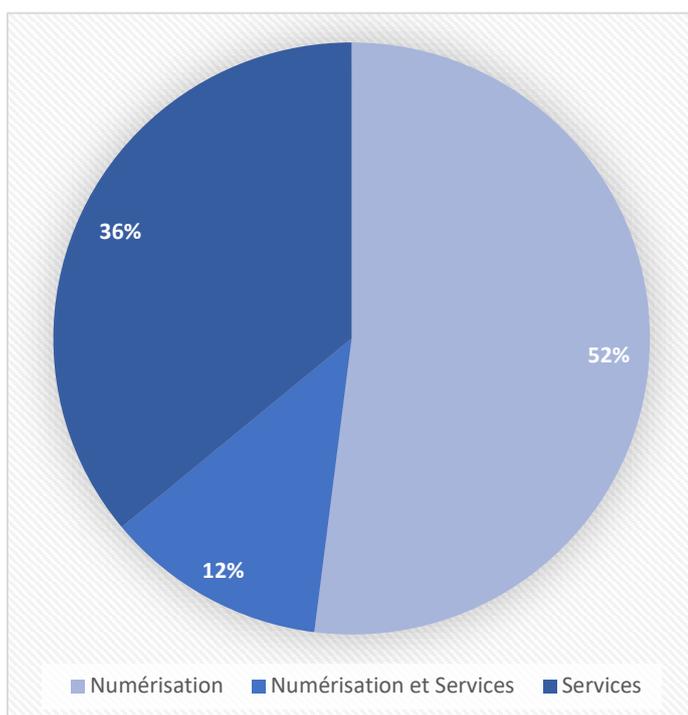


46 projets déposés (dont 1 inéligible)

Coût total : 8.037.748 €

Demande de subvention CollEx-Persée :
3.553.735 € = 44,2 % du coût total

Numérisation : 21
Numérisation et Services : 10
Services : 15



25 projets lauréats

Coût total : 4.356.080 €

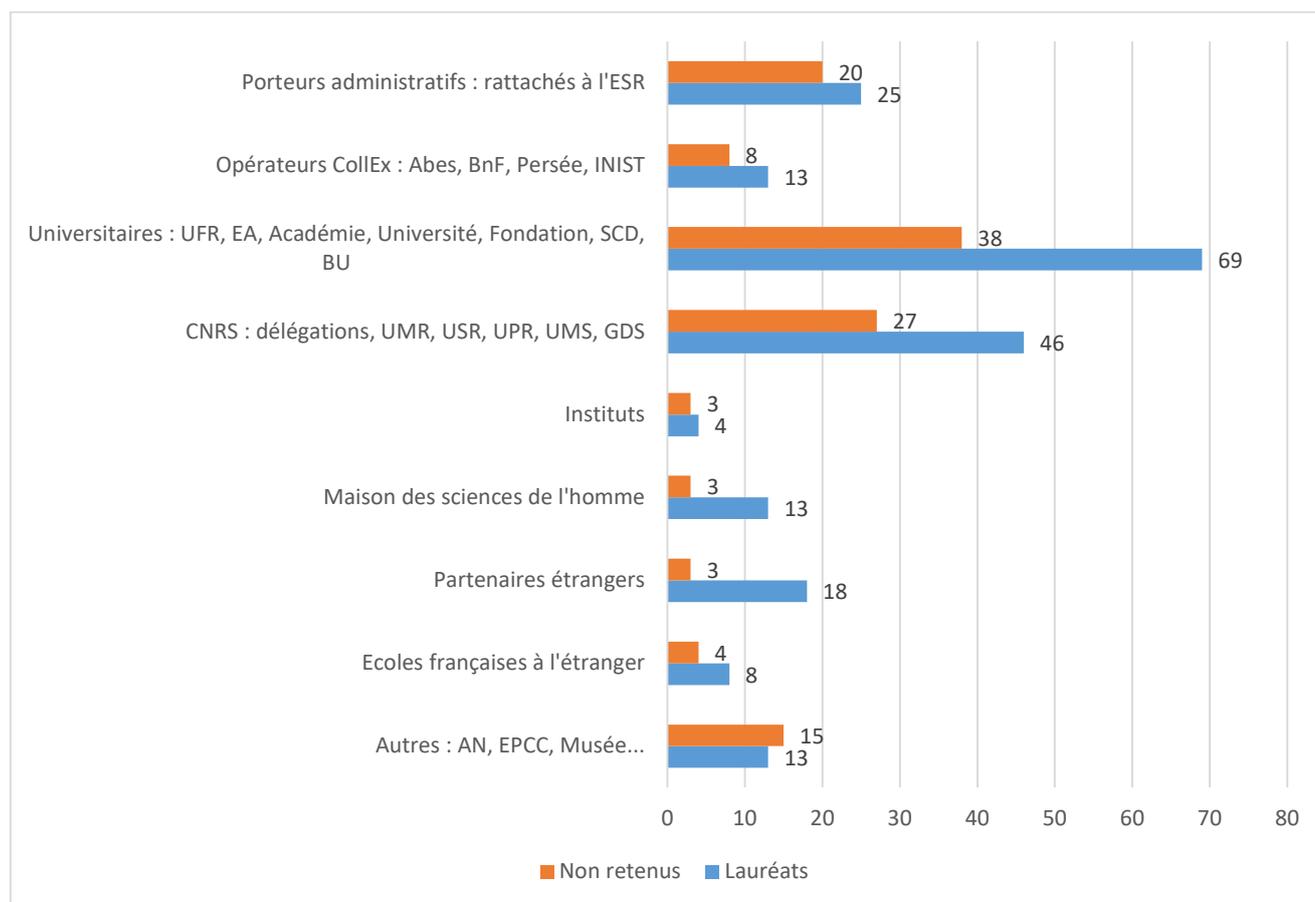
Subvention CollEx-Persée :
1.997.886 € = 46 % du coût total

Numérisation : 13
Numérisation et Services : 3
Services : 9

Remarque 1 : CollEx-Persée soutient financièrement à hauteur de 46% des projets co-construits IST/recherche, devenant ainsi un acteur essentiel du soutien à la recherche française, inscrit dans la science ouverte.

Remarque 2 : les typologies envisagées et choisies par les porteurs eux-mêmes au moment du dépôt seront à mutualiser. En effet, bien souvent, un projet numérisation comporte un axe services à la recherche, et inversement, rendant ainsi ces catégories beaucoup moins étanches que cela peut paraître.

Porteurs et partenaires engagés dans les projets

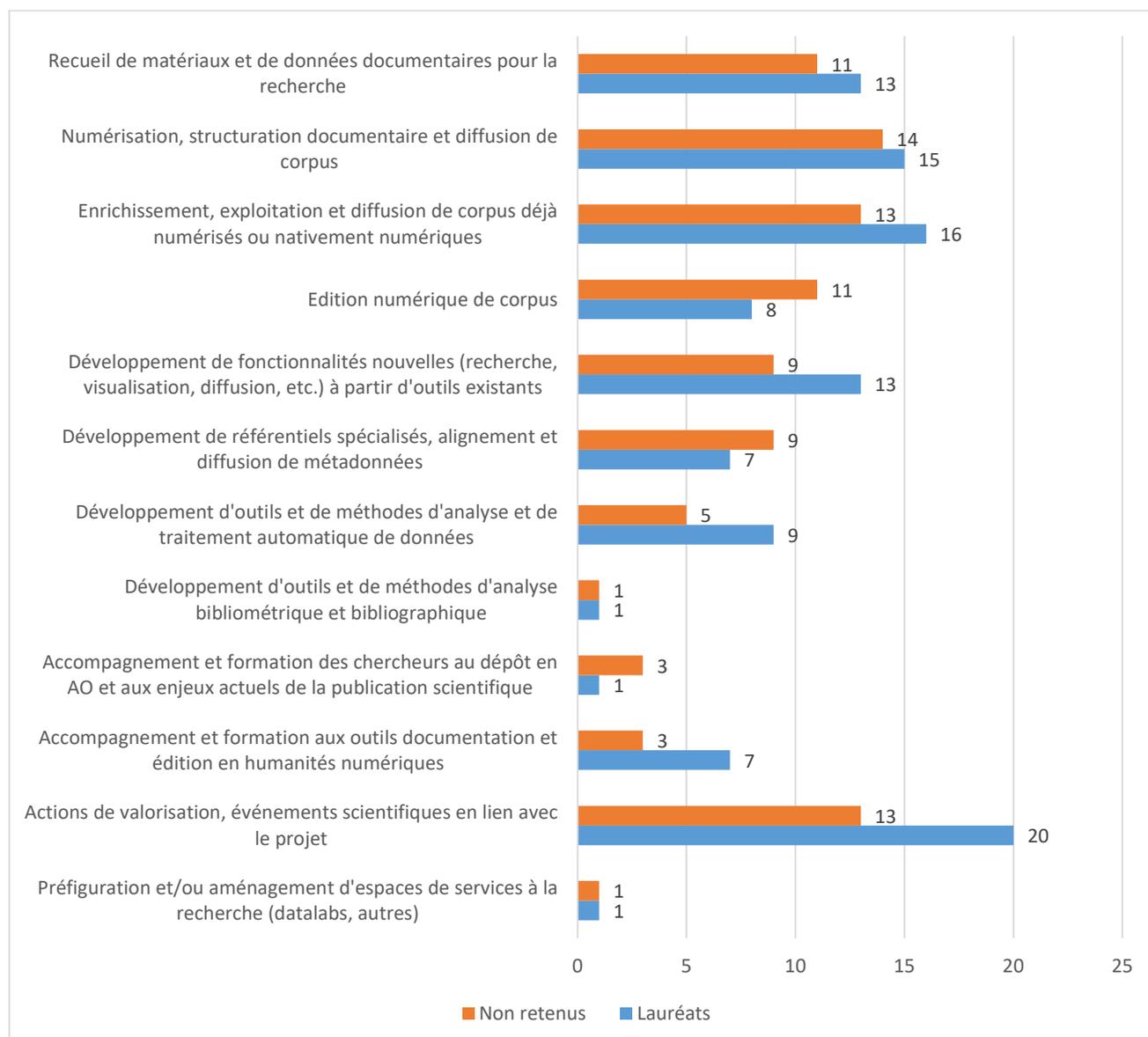


Remarque 1 : l'impact de l'appel à projets CollEx-Persée est très fort sur des communautés diverses, principalement ESR et CNRS. Au total, 330 porteurs et partenaires sont engagés dans les projets déposés. 209 parmi les lauréats, 121 parmi les non retenus. Certains d'entre eux sont en réalité présents dans plusieurs projets : l'Abes, la BnF, Persée, l'EFA, le RNMSH, l'Université de Poitiers, de Nanterre, de Bourgogne, le Campus Condorcet, le CIRAD, l'UMR 7366 (Georges Chevrier), le GDS Frantiq, etc...

Remarque 2 : les opérateurs CollEx et la BnF sont cités 21 fois et présents dans 18 projets, dont 13 lauréats.

Remarque 3 : les MSH se sont mobilisées pour cet appel à projets et se retrouvent présentes dans 7 projets lauréats, en particuliers celles de Poitiers, Dijon, Rennes, Aix-Marseille et Besançon. De même, le réseau des Ecoles françaises à l'étranger s'est mobilisé et ses membres sont présents dans 3 projets lauréats.

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'IST peut développer afin de répondre aux besoins de la recherche.

Remarque 1 : les points saillants tournent autour du recueil de matériaux et de données, de la constitution et de l'exploitation de corpus numériques, de l'édition numérique ainsi que du développement de fonctionnalités nouvelles d'outils existants. Les actions de valorisation liées aux projets constituent un axe très bien pris en compte par les porteurs.

Remarque 2 : trois sujets peuvent être considérés comme émergents fortement : celui lié aux référentiels, alignement et diffusion de métadonnées, le traitement automatique de données et la formation aux outils des humanités numériques.

Remarque 3 : trois sujets semblent peu couverts par la communauté CollEx-Persée : la création de Labs, l'accompagnement des chercheurs au dépôt en accès ouverts, les outils d'analyse bibliométriques. S'agit-il de sujets encore trop neufs, ou, au contraire, déjà bien développés dans d'autres cadres (au sein des SCD, de l'ADBU, du CoSO, etc.) ?

Disciplines représentées au sein des projets lauréats

D'un point de vue disciplinaire, cet appel à projets est 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.

De nombreux projets sont, en réalité, interdisciplinaires, avec une dominante HCERES :

Sciences humaines et sociales (80%) :

- *Monde ancien et contemporain* : histoire (8), histoire de l'art (1), archéologie (4) ;
- *Langues, textes, arts et cultures* : Langues / littératures anciennes et françaises, littérature comparée (2), Littératures et langues étrangères, civilisations, cultures et langues régionales (1) ;
- *Normes, institutions et comportements sociaux* : droit (1), ethnologie (2), sociologie (1) ;

Sciences et techniques, du vivant et de l'environnement (20%) :

- *Mathématiques* (1)
- *Physique/Chimie* (1)
- *Agronomie, biologie végétale, écologie, environnement, évolution* (3)

Annexe 1 : Liste des lauréats

ABRICO 2 - Archives, brochures et informations communistes ▪ Université de Bourgogne

Céline Alazard (CNRS - MSH de Dijon) et Jean Vigreux (Université de Bourgogne).

Acadie Sonore ▪ Université de Poitiers

Michaël Nauge (Université de Poitiers - MSHS) et André Magord (Université de Poitiers - MSHS).

AMOr – Archives Mathématiques d'Orsay ▪ Université Paris-Saclay

Elisabeth Kneller (Bibliothèque Jacques Hadamard – Saclay) et Emmylou Haffner (Université Paris-Saclay).

ArchéoRef Alignements ▪ Institut français d'archéologie orientale du Caire

Agnès Macquin (Ifao) et Laurent Coulon (Ifao).

ArchivesBaelo ▪ Casa de Velázquez

Elisabeth Collantes (CVZ - Bibliothèque et archives) et Laurent Callegarin (CVZ).

Archivos 3.0 ▪ Université de Poitiers

Rodolphe Defiolle (Université de Poitiers - MSHS Poitiers) et Fatiha Idmhand (Université de Poitiers).

BaOIA - Boîte à Outils d'Intelligence Artificielle ▪ Université Paris Nanterre

Cécile Tardy (La Contemporaine - Nanterre) et Julien Schuh (Université Paris Nanterre)

BibracteVilleOuverte ▪ Centre National de la Recherche Scientifique

Véronique Humbert (GDS Frantiq) et Philippe BARRAL (Université de Franche-Comté - MSHE).

Chronika ▪ Ecole française d'Athènes

Catherine Bouras (Efa) et Amélie Perrier (Efa).

CLA++ - Cabinet du livre d'artiste ▪ Université Rennes 2

Morgane Mignon (MSH en Bretagne) et Leszek Brogowski (Université Rennes 2).

CLIODIFLE – 1919-1985 : Une contribution à l'histoire de la francophonie ▪ Sorbonne Nouvelle Paris 3

Romain Le Nezet (Direction des BU – Université Sorbonne Nouvelle) et Valérie Spaëth (Université Sorbonne Nouvelle).

CONVEX 2 - Collection numérique vitivinicole d'excellence ▪ Université de Bourgogne

Agnès Viola (CNRS - MSH de Dijon) et Olivier Jacquet (Université de Bourgogne).

Datacc'+ - Dispositif d'accompagnement aux données de la recherche ▪ Université Claude Bernard Lyon 1

Marie-Emilia Herbet (SCD - UCB Lyon 1) et Thierry Billard (UMR5246 - UCB Lyon 1).

Datapoc 2.0 ▪ Muséum national d'histoire naturelle

Chloé Besombes (Direction des bibliothèques et de la documentation - MNHN) et Cécile Callou (MNHN - CNRS).

DISSINVENT - DISSidences de l'Est en exil : INVentaire, histoires, pratiques documentaires ▪ Université Paris Nanterre

Valérie Tesnière (La Contemporaine - Nanterre) et Sophie Cœuré (ICT - EA 337 Université de Paris).

ÈS LETTRES ▪ Université Paris 1 Panthéon-Sorbonne

François Haas (Bibliothèque interuniversitaire de la Sorbonne - UP1) et Pierre Verschueren (Université de Franche-Comté).

FamilieTTres ▪ Conservatoire national des arts et métiers

Cécile Formaglio (CNAM) et Michel Lallement (CNAM).

FemEnRev - Féminismes en revue ▪ Université d'Angers

Nathalie Clot (BU d'Angers), Magali Guaresi (Université Côte d'Azur) et Christine Bard (Université d'Angers).

ISSA - Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte ▪ Centre de coopération internationale en recherche agronomique pour le développement

Anne Toulet (DIST Cirad) et Franck Michel (Université Côte d'Azur - CNRS - Inria).

La guerre du droit ▪ Université Paris 1 Panthéon-Sorbonne

Alexandra Gottely (Bibliothèque interuniversitaire Cujas – Paris 1) et Florent Garnier (Université Toulouse 1 Capitole).

METADICTMEDICA - Metadictionnaire médical multilingue de la bibliothèque numérique Medica ▪ Université de Paris

Jean-François Vincent (BIU Santé - Université de Paris) et Nathalie Rousseau (CNRS - Sorbonne Université).

NumBois ▪ Centre de coopération internationale en recherche agronomique pour le développement

Sylvie Vago (DIST Cirad) et Jean Gérard (Unité de recherche BioWooEB du Cirad).

PALiBr - Pensée Anarchiste et Libertaire en Brochures ▪ Etablissement Public Campus Condorcet

Alicia León y Barella (GED Condorcet), Barbara Bonazzi (CHS Mondes contemporains) et Paul Boulland (CHS Mondes contemporains).

POPP - Projet d'Océrisation des recensements de la Population Parisienne ▪ Centre national de la recherche scientifique

François Merveille (GED Condorcet) et Sandra Brée (LAHRA - UMR5190).

Transcrire ▪ Musée du Quai Branly - Jacques Chirac

Paul-Emmanuel Bernard (MQB) et Anaïs Wion (Institut des mondes africains - UMR8171).

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.

1. ANDRO Mathieu, Institut national de la recherche agronomique
2. ANGEVELLE Catherine, Université de Lorraine – SCD
3. AYMOUNIN David, Agence bibliographique de l'enseignement supérieur
4. BERMES Emmanuelle, Bibliothèque nationale de France
5. BERTRAND Sophie, Bibliothèque nationale de France
6. BLIN Frédéric, Bibliothèque nationale et universitaire
7. BOISSAT Romain, Maison de l'Orient et de la Méditerranée Jean Pouilloux
8. BREUIL Sandrine, Université de Tours, Centre d'études supérieures de la Renaissance
9. CHAUDOREILLE Hélène, Université Paris 3 - BIU Sainte Geneviève/Bibliothèque Sainte Barbe
10. CRUGUEL Amanda, Université Lyon 1 - SCD
11. DAVRIEUX DE BECDELIEVRE Cécile, Bibliothèque municipale de Lyon
12. DE KOK Véronique, DISTRD-MESRI
13. DELESPIERRE Louis, DISTRD-MESRI
14. DELRUE Laure, Université de Lille - SCD
15. FORCADELL Sophie, Bibliothèque de Sciences Po Paris
16. FRANCOIS Claire, INIST-CNRS
17. GARCIA FERNANDEZ Anne, Université Grenoble Alpes, ELAN
18. GEROUDET Marie-Madeleine, Université de Lille - SCD
19. GIROLD Elise, Bibliothèque nationale et universitaire
20. GOTTELY Alexandra, Université Paris 1 Panthéon-Sorbonne - BIU Cujas
21. GRESLOU Elisabeth, Université Grenoble Alpes, Litt&Arts
22. GUICHARD Benjamin, BULAC
23. ILLIEN Gildas, Muséum national d'histoire naturelle
24. KLOPP Sophie, Bibliothèque Sciences Po Lyon
25. LAME-BERGIS Alix, Bibliothèque nationale de France
26. LAURENCEAU Amélia, Université Paris 1 Panthéon-Sorbonne - BIU de la Sorbonne
27. LE TELLIER BECQUART Nathalie, MSH Mondes
28. LEBRE Céline, Université de Nanterre - La Contemporaine
29. LORET Stéphane, Université de Nantes, ESO
30. MAGNIER Donatienne, Bibliothèque Sciences Po Paris
31. MARCEROU Nathalie, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques
32. MARLET Olivier, Université de Tours, UMR CITERES
33. MIURA Grégory, Université Bordeaux Montaigne - SCD
34. NIZIERS Guillaume, Centre technique du livre et de l'enseignement supérieur
35. OKRET-MANVILLE, Christine, Université Paris Dauphine - Bibliothèque
36. OURY Clément, Muséum national d'histoire naturelle
37. RAGEOT Laurence, Université de Tours, Maison des Sciences de l'Homme Val de Loire
38. RIVIÈRE Pauline, Université Sorbonne Nouvelle Paris 3, BIU Sainte-Geneviève
39. ROSSI Pier Luigi, Institut de recherche pour le développement
40. SCHERER Marc, Université Paris 3 - BIU Sainte Geneviève
41. SCHWEITZER Jérôme, Bibliothèque nationale et universitaire
42. STOLL Mathieu, MCC-Service Interministériel des Archives de France
43. STOLL Véronique, Observatoire de Paris - Bibliothèque
44. TARDY Cécile, Université Paris Nanterre - La Contemporaine
45. TISSEAU DES ESCOTAIS Aleth, Observatoire de Paris - Bibliothèque
46. WENZ Romain, Université de Bordeaux – SCD
47. WESTEEL Isabelle, Bibliothèque municipale de Grenoble

Annexe 3 : Résumés en français et en anglais des projets retenus

ABRICO 2

Archives, brochures et informations communistes

Université de Bourgogne

Céline Alazard (CNRS – MSH de Dijon) et **Jean Vigreux** (Université de Bourgogne)

Le projet ABRICO 2 (Archives, brochures et informations communistes) se donne pour objectif la numérisation et la mise en ligne de documents politiques relatifs au mouvement communiste international. Porté par la MSH de Dijon, labellisée pour sa collection « Critique et Mouvements sociaux », ABRICO 2 se veut le prolongement de précédents travaux menés dans le but de favoriser l'accessibilité des citoyens et des chercheurs à un ensemble de documents communistes liés au second XXème siècle. La documentation privilégiée par ce projet se caractérise par l'interdisciplinarité (philosophie, politologie, histoire, littérature, etc.) et une identité aussi bien nationale qu'internationale, reflétant ainsi le caractère global de l'idée communiste. C'est pourquoi nous nous attachons à associer des partenaires étrangers à notre projet, notamment l'Université Manouba de Tunis et l'Université Roma 1. De plus, l'approche du centenaire du Parti communiste français offrira un relief particulier à nos travaux en les intégrant à l'actualité historique.

English summary : The project ABRICO 2 (Archive, brochures and communists informations) has the will to digitize and upload several political documents related to the international communist movement. Lead by the Maison des Sciences de l'Homme de Dijon, recently labeled for its collection « Criticism and social movements », ABRICO 2 wants to pursue former works led in order to promote a better accessibility to a set of communist documents tied to the second twentieth century, especially for citizens and reseachers. The project puts forward a documentation characterized by interdisciplinarity (philosophy, politology, history, literature, etc.) and a national/international identity which exemplifies the global identity of the communist idea. That is the reason why we strive to associate international partners to our project, including the University of Manouba (Tunis) and the University Roma 1. Moreover, the future centenary of the French communist party's foundation will resonate with our works, by integrating them into an historical agenda.

Acadie Sonore

Université de Poitiers

Michaël Nauge (Université de Poitiers - MSHS) et **André Magord** (Université de Poitiers - MSHS)

« Acadie sonore » s'insère dans un projet plus vaste de documentation et de valorisation via une plateforme numérique (« Francoralité ») d'archives sonores collectées dans le Grand-Ouest français et en Amérique du Nord francophone. Une étape préliminaire du projet « Acadie sonore » a permis de le positionner au sein des actions portées par la MSHS de Poitiers et de l'intégrer aux travaux de son équipe informatique. Aujourd'hui, ces informaticiens nous épaulent dans les liens aux développeurs de la plateforme numérique sur laquelle sont déposés les fonds et aux équipes d'Huma-Num où sont hébergées les données. Nous sommes actuellement à un moment charnière avec d'un côté, un important corpus audio préalablement collecté avec l'aide de nos partenaires et, de l'autre, une plateforme web collaborative pour le rayonnement de celui-ci. Il nous faut maintenant intégrer massivement ce corpus dans la plateforme.

Nous pouvons identifier deux grandes actions à mener en parallèle:

Action 1. Corpus local : travail de restructuration/régularisation du corpus et de ses descriptions pour un dépôt massif. Nous souhaitons apporter une attention particulière à la qualité des métadonnées structurées en Dublin Core Dcterms associé à une curation et validation par l'utilisation de "goodtables" recommandé par data.gouv.

Action 2. Plateforme web : augmenter le respect des paradigmes FAIR, avec une utilisation avancée des services d'Huma-Num suivants: Isidore, Nakala et Opentheo. C'est un travail que nous menons de manière coordonnée et partagée avec le consortium Archive Des Ethnologues.

English summary : "Acadie Sonore" is part of a larger project of documentation and promotion of sound archives collected in the French Great West and in French-speaking North America. A preliminary stage of the "Sound Academy" project made it possible to position it within the actions carried out by the MSHS of Poitiers and to integrate it into the work of its IT team. Today, these computer scientists support us in the links to the developers of the digital platform (Francoralité) on which the funds are deposited and to the Huma-Num teams where the data are hosted.

Action 1. Local Corpus : work of restructuring / regularization of the corpus and its descriptions for a massive deposit. We would like to pay particular attention to the quality of metadata structured in Dublin Core Dcterms associated with curation and validation by the use of "goodtables" recommended by data.gouv.

Action 2. Web platform : increase respect for FAIR paradigms, with advanced use of the following Huma-Num services : Isidore, Nakala and Opentheo. This is work that we are doing in a coordinated and shared manner with the Archive of Ethnologists consortium.

Archives Mathématiques d'Orsay

Université Paris Saclay

Elisabeth Kneller (Bibliothèque Jacques Hadamard - Saclay) et **Emmylou Haffner** (Université Paris-Saclay)

Le projet Archives Mathématiques d'Orsay (AMOr) vise à développer les stratégies d'éditorialisation scientifique et de patrimonialisation savante des archives de la recherche mathématique. Pour cela, nous travaillerons sur les fonds d'archives mathématiques du Laboratoire de Mathématiques d'Orsay et de l'Institut des Hautes Études Scientifiques (datant de 1950 à 1990), qui concernent la recherche fondamentale, la vie académique et l'enseignement supérieur en mathématiques. Ce projet interdisciplinaire et transversal rassemble plusieurs bibliothèques et équipes de l'Université Paris-Saclay et de l'ENS Ulm. Il se donne pour buts :

1. Collecter, trier et cataloguer les archives de la recherche et de l'enseignement mathématiques à Orsay. Identifier les parties du corpus éligibles à la numérisation et commencer à défricher ce travail de numérisation. Mettre en place des stratégies d'archivage des archives nativement numériques.
2. Créer des modèles d'apprentissage pour la reconnaissance de caractères (OCR et HTR) en mathématiques capable de reconnaître d'une équation et la transformer en code LaTeX ou MathML, produire des fichiers structurés et encodés en XML et LaTeX.
3. Une édition numérique des archives (fac-similés et transcriptions) avec Omeka.
4. Développer de nouvelles recherches en histoire des mathématiques et de l'enseignement.

English summary : The "Archives Mathématiques d'Orsay" (AMOr, Mathematical Archives in Orsay) project aims at developing new strategies for the editorialization and academic patrimonialization of the archives of mathematical research. To do so, we will work on archives from the Laboratoire de Mathématiques d'Orsay and the Institut des Hautes Études Scientifiques (dating from 1950 to 1990). These documents deal with fundamental research, academic life and administration, and higher education in mathematics. This interdisciplinary and transversal project brings together several libraries and research teams from the Université Paris-Saclay and ENS Ulm. Its goals are:

1. Collect, sort and catalog archives of mathematical research and teaching in Orsay. Identify which parts of the corpus could be digitized, and start to clear paths for this part of the work. Set up archiving strategies of natively digital archives.
2. Create training models for Optical Character Recognition and Handwritten Text Recognition for mathematics, which would be able to recognize an equation and transform it into LaTeX or MathML code, to produce structured and encoded files in XML and LaTeX.
3. A digital edition of the archives (facsimiles and transcriptions) using Omeka.
4. Develop new research in the history of mathematics and the history of teaching.

ArchéoRef Alignements

Institut français d'archéologie orientale du Caire
Agnès Macquin (Ifao) et **Laurent Coulon** (Ifao)

Comme suite à la subvention obtenue par les cinq Ecoles françaises à l'étranger en 2015 sur programme ISTEK pour le projet ArchéoRef, les EFE souhaitent compléter les données concernant des sites et éléments de site de leurs chantiers archéologiques pour un enrichissement IdRef noms géographiques. Le présent projet s'accompagne d'un alignement avec PACTOLS et, pour les données de géolocalisation, d'alignements avec Geonames et Pleiades, grâce à la collaboration de la BIS et du GDS Frantiq.

English summary : Following the grant obtained by the five French Schools abroad in 2015 on the ISTEK programme for the project ArcheoRef, the EFEs propose to complete the data regarding the sites and site elements of their archaeological sites for an IdRef geographical names enrichment. The current program is accompanied by an alignment with PACTOLS and, for geolocation data, alignment with Geonames and Pleiades, thanks to the collaboration of BIS and the GDS Frantiq.

ArchivesBaelo

Casa de Velázquez

Elisabeth Collantes (CVZ – Bibliothèque et archives) et **Laurent Callegarin** (CVZ)

Bénéficiant d'une approche interdisciplinaire et d'une dimension internationale, le projet ARCHIVESBAELO a pour objet la conservation, l'archivage, la numérisation et la valorisation de la documentation archéologique française produite entre 1917 et 2017 par les chercheurs liés à la Casa de Velázquez (Madrid), sur le site antique de Baelo (Tarifa, Cadix). L'idée est de regrouper les documents aujourd'hui dispersés dans différents fonds sur une même base numérique, afin de constituer un réservoir de données scientifiques. Les ensembles d'archives conservés à la Casa de Velázquez et à l'Institut de recherche sur l'architecture antique (Pau) sont constitués d'une grande variété d'éléments papier qui représentent huit mètres linéaires. Les tâches sont : intégrer les nouveaux versements, poursuivre et finaliser la numérisation des fonds détenus par des institutions françaises, récupérer et mettre en forme les métadonnées des institutions espagnoles, déposer les fichiers numérisés dans Nakala, mettre à disposition des chercheurs mais également du public un instrument de recherche sous AtoM et parallèlement développer une plateforme numérique bilingue permettant d'éditorialiser l'ensemble de la documentation archéologique du site de Baelo. A l'issue du programme (à l'horizon 2022), l'objectif est d'intégrer ce corpus dans l'espace « chantiers de fouilles » de la future Perséide des Écoles françaises à l'étranger.

English summary : Benefiting from an interdisciplinary approach and an international dimension, the ARCHIVESBAELO project aims to preserve, archive, digitize and enhance the French archaeological documentation produced between 1917 and 2017 by researchers associated with the Casa de Velázquez (Madrid), on the ancient site of Baelo (Tarifa, Cadiz). The idea is to group together the documents now scattered in different collections on the same digital basis, in order to constitute a reservoir of scientific data. The archival collections kept at Casa de Velázquez and the Research Institute for Ancient Architecture (Pau) consist of a wide variety of paper elements representing eight linear metres. The tasks are : to integrate the new payments, to continue and finalize the digitization of the funds held by French institutions, to recover and format the metadata of Spanish institutions, to deposit the digitized files in Nakala, to make available to researchers but also to the public a research tool under AtoM and at the same time to develop a bilingual digital platform allowing the editorialization of all the archaeological documentation of the Baelo site. At the end of the programme (by 2022), the objective is to integrate this corpus into the "excavation sites" of the future Perseide of French schools abroad.

Archivos 3.0

Université de Poitiers

Rodolphe Defiolle (Université de Poitiers – MSHS Poitiers) et **Fatiha Idmhand** (Université de Poitiers)

Le projet Archivos 3.0 est un projet trans-disciplinaire qui porte sur les fonds d'archives de figures majeures de la littérature et de la culture latino-américaine. Ils sont traités et étudiés depuis la littérature, la génétique des œuvres, l'histoire, les études culturelles, l'archivistique et l'informatique. Les fonds concernés permettent d'aborder le cas d'intellectuels et écrivains latino-américains depuis la Caraïbe jusqu'au Río de la Plata dont les manuscrits, articles, correspondances et documents apportent des éclairages importants sur les moyens de la circulation, de la réception et de la création de la littérature latino-américaine durant les périodes de conflits politiques des années 1970 et de crises économiques des années 2000. Archivos 3.0 comporte deux volets qui évoluent de façons complémentaires : une partie du programme est dédiée à la reprise de données numériques existantes, l'autre concerne la création de données numériques inédites. Les enjeux de l'un et l'autre en termes de qualité technique et d'organisation conceptuelle des données se rejoignent dans le but de préparer un corpus exploitables dans l'environnement du web de données : le web 3.0.

English summary : Archivos 3.0 is a trans-disciplinary project that focuses on the archival funds of major Latin American literature and culture figures. They are studied from the fields of literature, genetics and creative processes, history, cultural studies, archival sciences and computer sciences. The collections concerned address the case of Latin American intellectuals and writers from the Caribbean to the Río de la Plata, whose manuscripts, articles, correspondence and documents highlight ways of circulation, reception and creation of Latin American literature during the periods of political conflict in the 1970s and the economic crises of the 2000s. Archivos 3.0 has two components that evolve in complementary ways : a part of the program is dedicated to the recovery of existing digital data, the other to the creation of new digital data. The stakes of both in terms of technical quality and conceptual organization of the data come together with the aim of preparing a corpus that can be exploited in the web of data environment : web 3.0..

Boîte à Outils d'Intelligence Artificielle

Université Paris Nanterre

Cécile Tardy (La Contemporaine - Nanterre) et **Julien Schuh** (Université Paris Nanterre)

Ce projet a pour objectif d'offrir une boîte à outils d'analyse de données massives pour des projets consacrés aux sciences humaines et sociales appuyés sur des archives et des collections numérisées. L'idée centrale est de développer des compétences pour les chercheurs et les professionnels de la documentation par l'initiation au Machine Learning et aux outils pour traiter différentes archives (affiches, guides, périodiques, estampes) et de fournir des outils automatisés de traitement et de valorisation de ces corpus (automatisation de l'étiquetage, plug-ins pour des plateformes de consultation comme Omeka). La boîte à outils reposera sur des modèles pouvant être aisément adaptés à différents usages et comportera des programmes réutilisables (classification d'images, d'unités textuelles, reconnaissance de structures, détection de similarités...). Plusieurs sous-projets à valeur de modélisation conduisant à la réalisation d'outils et permettant d'affiner les modèles seront mis en œuvre.

English summary : The objective of this project is to provide a massive data analysis toolkit for projects dedicated to the Human and Social Sciences based on digitized archives and collections. The central idea is to develop skills for researchers and curators by introducing them to Machine Learning and tools for processing different archives (posters, guides, periodicals, prints) and to provide automated tools for processing and valorizing these corpora (automation of labeling, plug-ins for consultation platforms such as Omeka). The toolbox will be based on models that can be easily adapted to different uses and will include reusable programs (classification of images, text units, structure recognition, detection of similarities, etc.). Several showcase sub-projects leading to the creation of tools and allowing the refinement of the models will be undertaken.

BibracteVilleOuvverte

Centre National de la Recherche Scientifique

Véronique Humbert (GDS Frantiq) et **Philippe Barral** (Université de Franche-Comté - MSHE)

«BibracteVilleOuvverte» associe deux volets du « cycle de vie » de la connaissance archéologique autour d'une même problématique de recherche : le partage et la diffusion d'une collection d'ouvrages, de données numériques et d'archives numérisées d'une part, la gestion et le partage du catalogue d'une bibliothèque spécialisée d'autre part.

- Une bibliothèque scientifique virtuelle sur un site archéologique emblématique de la fin de l'âge du Fer réunira une documentation variée : plans, dessins, carnets de fouille, publications, données de la recherche, nativement numérique ou numérisée ;

- Les ± 70.000 références du catalogue de la bibliothèque de Bibracte, qui fait autorité pour la protohistoire européenne, seront documentées au moyen des PACTOLS enrichis par un thésaurus *ad hoc* – cf ci-dessous – et seront versées dans le catalogue collectif indexé du GDS Frantiq.

Les deux volets du projet partageront un thésaurus commun (vocabulaire scientifique et données d'autorité), créé sur Opentheso et aligné avec les grands référentiels du web sémantique ; ce « référentiel de l'âge du Fer » francophone enrichira PACTOLS, ensemble de vocabulaires contrôlés porté par le GDS Frantiq.

Une partie des archives concernées aura fait l'objet d'une transcription et d'une documentarisation participatives associant professionnels de l'IST, chercheurs archéologues et amateurs. Le corpus virtuel ainsi constitué a vocation à enrichir la « Perséide » *Bibracte* qui sera diffusée par l'UMS Persée et connectée au web sémantique via le *triple store* associé.

English summary : "BibracteVilleOuvverte" combines two steps of the archaeological knowledge "life cycle" on the same research issue : the sharing and dissemination of a collection of books, digital data and digitized archives on the one hand, and the management and sharing of the catalogue of a specialised library on the otherhand.

- A virtual scientific library on an emblematic archaeological site from the end of the Iron Age will bring together a variety of documentation : maps, drawings, excavation notebooks, publications, research data, natively digital or digitized ;

- The ± 70.000 references in the Bibracte library catalogue, which is an authoritative catalogue for European protohistory, will be documented by means of the PACTOLS enriched by an *ad hoc* thesaurus - see below - and will be included in the "catalogue collectif indexé" (CCI) of the GDS Frantiq.

The two parts of the project will share a common thesaurus (scientific vocabulary and authority data), created within Opentheso and aligned with the major repositories of the semantic web ; this French-speaking "Iron Age repository" will enrich PACTOLS, a set of controlled vocabularies carried by the GDS Frantiq.

Part of these archives have been transcribed and documentarized in a participatory process involving IST professionals, archaeologists and amateurs. The virtual corpus thus compiled is to enhance the Bibracte "Perseid" which will be made available by the UMS Persée and connected to the semantic web by means of the associated triple store.

Chronika

Ecole française d'Athènes

Catherine Bouras (EfA) et **Amélie Perrier** (EfA)

Depuis 100 ans, la Chronique est un outil de travail indispensable dans le domaine de l'archéologie de la Grèce. Depuis sa première parution en 1920, son objectif est d'offrir à la communauté scientifique une vue d'ensemble des travaux archéologiques réalisés chaque année. Paraissant d'abord dans le Bulletin de Correspondance Hellénique, elle s'est profondément transformée au cours de ces dernières années. La Chronique des fouilles est devenue en 2009 une base de données en ligne dotée d'un moteur de recherche puissant qui combine tous les champs disponibles de la base afin de permettre une recherche la plus précise possible sur les fouilles réalisées chaque année. Cette base de données s'inscrit dans un partenariat international en associant la British School at Athens et sa publication *Archaeological Reports*.

L'objectif du projet « Chronika » est de compléter la base de données en réunissant la totalité de la documentation relative à la Chronique des fouilles en Grèce depuis ses débuts. Ce travail consiste en la mise en ligne rétrospective de 84 années de fouilles publiées dans le BCH et consultables sur le portail Persée. Cette reprise des données consiste à produire, grâce à la chaîne de traitement de Persée, une structure documentaire d'une granularité plus fine conformément à celle attendue par la base de la Chronique. Ensuite, elle se poursuivra à Athènes par la vérification et correction des informations (issues de l'OCR), l'attribution d'un DOI, l'alignement aux référentiels et par un lien éventuel avec des informations complémentaires issues des archives de l'EfA (archimages).

Cet enrichissement sera alors accessible sur le portail de la chronique des fouilles mais également moissonnable (protocole OAI-PMH) ou interrogeable via des requêtes Sparql.

A terme Persée réintégrera cette plus-value documentaire au sein de son portail. Ce travail enrichira l'espace « fouilles archéologiques » de la future Perséide « Écoles françaises à l'étranger ».

English summary : The "Chronique des fouilles" was first published in 1920 as a part of Bulletin de Correspondance Hellénique (BCH), by the French School at Athens. Its aim is to offer to the reader an overall view of excavations and archaeological discoveries throughout Greece each year. For 100 years now, the Chronique des fouilles has been an important research tool for researchers, teachers, students and informed readers.

As the review had to face a growing documentation and the published volume each year became larger, it was decided to transform the printed article into an online database that would have a full search engine, combining all fields in order to allow the most precise research. This was decided as an international partnership with the British School at Athens and its publication *Archaeological Reports*.

The objective of the "Chronika" project is to bring together in a single database all the documentation relating to the Chronique des fouilles en Grèce/Archaeology in Greece Online. This work therefore consists in the retrospective online publication of 84 years of archaeological reports published by the BCH and available for consultation on the Perseus portal. This data recovery will be based on an initial documentary work by the Persée team and will then be continued in Athens by checking and correcting the information (from the OCR), assigning a DOI, aligning with chronological (PeriodO), geographical (Geonames, Pleiades) and subject (Pactols) repositories and possibly linking with additional information from the EfA archives (archimages).

Cabinet du livre d'artiste

Université Rennes 2

Morgane Mignon (MSH en Bretagne) et **Leszek Brogowski** (Université Rennes 2)

Après la labellisation CollEx du fonds du Cabinet du livre d'artiste (CLA) en 2018, le projet CLA++ constitue la première étape vers une conversion du fonds. En tant que ressources documentaires spécifiques, les publications d'artistes nécessitent un traitement particulier pour leur numérisation et leur catalogage. Si ces opérations s'appuient sur des règles communes définies en amont, la prise en compte du double statut d'objets de recherche et d'œuvres d'art est cruciale pour réussir leur signalement au sein des réseaux de la recherche universitaire et des collections d'art.

Le projet se déploie en trois étapes sur 24 mois :

1. Constitution d'un répertoire des typologies de publications d'artistes et expérimentation de solutions techniques pour une numérisation adaptée (essais menés sur une sélection de 200 titres).
2. Amélioration de la base de données existante couplée à l'élaboration d'un guide de numérisation et d'un référentiel de catalogage pour les publications d'artistes, à la lumière d'expériences déjà menées dans des institutions académiques et/ou muséales et des nouvelles recommandations dans le contexte du Plan national pour la Science Ouverte.
3. Étude de la faisabilité technique d'une interopérabilité de la base de données du CLA avec le Sudoc et Videomuseum ; par extension, réflexion sur une mutualisation de ces ressources.

English summary : After the CollEx certification for the Cabinet du livre d'artiste (CLA) in 2018, the CLA++ project is the first step towards a conversion of the collection. As specific documentary resources, artists' publications require a special treatment for their digitization and cataloging. If these operations are based on common rules defined upstream, taking into account their dual status of research objects and works of art is crucial to successfully report them within university research networks and art collections. The project is deployed in three stages over 24 months :

1. Constitution of a directory of typologies of artists' publications and experimentation of technical solutions for an adapted digitization (tests conducted on a selection of 200 titles)
2. Improvement of the existing database along with the development of a digitization guide and cataloging reference system for artists' publications, in light of experiences already carried out in academic and/or museum institutions and of new recommendations in the context of the National Open Science Plan.
3. Study of the technical feasibility of interoperability of the CLA database with Sudoc and Videomuseum ; by extension, thinking about pooling these resources.

CLIODIFLE

1919-1985 : Une contribution à l'histoire de la francophonie

Sorbonne Nouvelle Paris 3

Romain Le Nezet (Direction des BU – Université Sorbonne Nouvelle) et **Valérie Spaëth**
(Université Sorbonne Nouvelle)

Le projet vise à numériser et mettre en ligne, de manière critique et enrichie, les fonds d'archives de deux institutions à l'origine de la didactique du français langue étrangère (FLE) : les Cours de Civilisation Française de la Sorbonne (CCFS) qui délivraient essentiellement, à l'origine, des cours de littérature, de civilisation et de langue françaises aux étrangers et l'Ecole de Préparation des Professeurs de Français à l'Etranger (EPPFE), qui formait les enseignants, en grande partie étrangers, à l'enseignement du FLE. Elles sont respectivement créées dans l'Académie de Paris, en Sorbonne, en 1919 et 1920.

Ces sources premières inédites contribueront au renouvellement de l'histoire de la didactique du FLE et de la construction de la francophonie scolaire et universitaire dans la mesure où de nombreuses nationalités de futurs professeurs de français sont impliquées. Ce corpus dépasse le seul périmètre de la Sorbonne. Il sera largement accessible à deux niveaux (formation et recherche) et sera valorisé de différentes manières : création d'une bibliothèque numérique, mise en place d'outils de recherche en ligne, et organisation d'événements (colloque, publications, conférences, exposition, etc.).

English summary : The project aims to digitalize and publish online a critical and enhanced version of the archival holdings of two main institutions within applied french linguistics. Les cours de Civilisation Française de la Sorbonne (CCFS) was, indeed, originally a school in charge of teaching french language, culture and literature to foreigners while l'Ecole de Préparation des Professeurs de français à l'Etranger (EPPFE) trained French and Foreigners students to teach French. The set wo institutions are both created inside Paris' Academy in the Sorbonne as early as 1919 for the CCFS and 1920 for the EPPFE. Those unpublished documents will contribute to the renewal of french applied linguistic's history and the construction of a french speaking scholarship context. The corpus of documents will be available for both training and research, therefore, it will be reachable online (through the libraries collections and search engines) so as in scientific events (conferences, publications, exhibitions...).

CONVEX 2

Collection numérique vitivinicole d'excellence

Université de Bourgogne

Agnès Viola (Université de Bourgogne - MSH de Dijon) et **Olivier Jacquet** (Université de Bourgogne)

L'objectif recherché est dans un premier temps de poursuivre la politique documentaire 3.0 initiée par la MSH de Dijon depuis plusieurs années, et ce grâce à la diffusion en accès ouvert et gratuit d'une revue professionnelle ancienne, inédite au format numérique : *Le Progrès agricole et viticole*. La seconde partie du projet consiste en l'élaboration d'un corpus numérique signalant des publications anciennes, du XVIIIe à 1950, déjà présentes en ligne ou restant à numériser. Le traitement de ces deux corpus, distinct en matière de méthodologie, favorise la mise en place d'approches holistiques et multiscalaires d'un intérêt scientifique indéniable. En outre, l'histoire des sciences et techniques vitivinicoles n'est pas la seule à profiter de ce projet puisque ce dernier englobe dans le même souci d'exhaustivité certains aspects juridiques, économiques et socio-culturels inhérents au monde de la vigne et du vin. L'objectif affiché permettra non seulement de mieux appréhender la place de la France dans le monde en tant que pays producteur, mais aussi d'envisager ses interactions avec d'autres nations viticoles au cours des trois derniers siècles. Enfin en se plaçant sous le signe de l'histoire globale, CONVEX 2 ne s'adresse pas uniquement aux chercheurs en sciences humaines et sociales, mais aussi aux professionnels de la filière vitivinicole ainsi qu'à un public curieux de découvrir un monde en constante mutation.

Ces ressources viendront alimenter les collections relatives à la vigne et au vin diffusées par la MSH de Dijon et labellisées par CollEx en 2017.

English summary : The aim at first is to pursue the 3.0 documentary policy initiated by the MSH de Dijon for several years, by uploading in free and open access an old professional journal, still unpublished in digital format : *Le Progrès agricole et viticole*. The second part consisting in the development of a digital corpus of ancient works, from the 18th century to 1950, already available online or still to be digitized. The handling of these two corpus, which is quite distinct in terms of methodology, favours the implementation of holistic and multiscalal approaches, both of an undeniable scientific interest. In addition, the history of wine science and technology is not the only one to benefit from this project, since it includes in the same concern of completeness, legal, economic and socio-cultural aspects related to the world of vine and wine. Our stated goal will allow not only a better understanding of France's place in the world as a producing country, but also a better grasp of its interactions with other wine-producing nations during the last three centuries. Finally, by placing itself under the sign of global history, CONVEX 2 is not only addressed to humanities and social sciences researchers but also to the professionals of the wine sector, as well as to an audience interested in discovering a world in constant change. Those documentary resources will feed our "vines and wines" collections, labeled CollEx-Persée since 2017

Datacc'+

Dispositif d'accompagnement aux données de la recherche

Université Claude Bernard Lyon 1

Marie-Emilia Herbert (SCD - UCB Lyon 1) et **Thierry Billard** (UMR5246 - UCB Lyon 1)

Datacc'+ propose un dispositif complet d'accompagnement des chercheurs en matière de gestion des données, depuis la production et la collecte jusqu'à la publication et la préservation de celles-ci. Le projet repose sur une approche originale, volontairement disciplinaire et conçue en adéquation avec les principes de la science ouverte. L'une des particularités du dispositif consiste à améliorer la traçabilité des travaux scientifiques, leur reproductibilité et l'efficacité de la recherche à travers la publication de résultats négatifs et le déploiement d'un cahier de laboratoire électronique

English summary : Datacc'+ proposes a global system aimed at assisting researchers in their data management, covering the production, collection, publication and conservation of information. The project is grounded in an original disciplinary approach and designed in accordance to open science methods. One of the distinct characteristics of the system includes the traceability of scientific research as well as its replication and optimization as it facilitates the publication of negative results and use of laboratory notebooks.

Datapoc 2.0

Muséum national d'histoire naturelle

Chloé Besombes (Direction des bibliothèques et de la documentation - MNHN) et **Cécile Callou** (MNHN - CNRS)

La preuve de concept datapoc.mnhn.fr, lauréate de l'AAP 2018, utilise les technologies du web sémantique pour explorer les possibilités d'alignement, de liage et de publication de données utiles à la recherche taxonomique (métadonnées, numérisations...) issues de plusieurs réservoirs de données du Muséum associées à un corpus d'environ 500 noms de savants et collecteurs naturalistes. Le prototype <https://datapoc.mnhn.fr> permet d'évaluer la pertinence d'un nouveau mode de navigation dans ces données liées. Il offre aussi des possibilités de réutilisation, notamment aux fins d'enrichissement d'agrégateurs et de référentiels. Les premiers résultats incitent à poursuivre le projet dans trois directions :

1. Dans la perspective d'un passage à l'échelle, l'extension du volume de données : inventaires d'archives, données d'observation, publications déposées dans HAL... Le corpus sera enrichi selon les priorités des chercheurs embarqués dans le projet.
2. Les traitements algorithmiques déployés ont révélé des problèmes de qualité et de structuration dans les données sources. Ce constat motive l'intention de doter le service d'un outil collaboratif de repérage et de correction des anomalies par les chercheurs et les collecteurs. Ces fonctionnalités seraient destinées tant aux scientifiques et gestionnaires du MNHN qu'aux amateurs, dans la dynamique des sciences participatives.
3. Il s'agira enfin de tester plus avant les mécanismes de réutilisation et d'enrichissement croisé des données en coopération avec des partenaires extérieurs : Fichier national d'entités (Abes/BnF), bibliothèque numérique BHL (Biodiversity Heritage Library), infrastructure européenne DISSCO (Distributed System of Scientific Collections).

English summary : The datapocproof of concept, winner of the 2018 AAP, uses semantic web technologies to explore the possibilities of aligning, linking and publishing data useful for taxonomic research (metadata, digitizations...) from several MNHN data repositories and catalogues associated with a corpus of 500 names of naturalist scientists. The <https://datapoc.mnhn.fr> prototype allows to assess the relevance of a new mode of navigation in this linked data. It also offers reuse opportunities, particularly for the purpose of enriching other national or international aggregators and repositories. The initial results suggest that the project should continue in three directions :

1. To prepare scaling and industrialization scenarios, the extension of the scope and volume of data : archival inventories, observational data, publications filed in the HAL French Open Access repository... The corpus will be enriched according to the priorities of the researchers involved in the project.
2. Algorithmic treatments deployed revealed quality and structuration issues in the source data. This finding motivates the intention to equip the service with a collaborative tool for identifying and correcting anomalies. These features would be intended for both scientists and curators of the MNHN and online amateurs, in the dynamics of citizen science.
3. Finally, we plan further testing of data reuse and cross-enrichment mechanisms in cooperation with external partners : National Entity File (Abes/BnF), BHL Digital Library (Biodiversity Heritage Library), European DISSCO (Distributed System of Scientific Collections) infrastructure.

DISSIVENT

DISSidences de l'Est en exil : INVENTaire, histoires, pratiques documentaires

Université Paris Nanterre

Valérie Tesnière (La Contemporaine - Nanterre) et **Sophie Cœuré** (ICT – EA 337 Université de Paris)

Le projet DISSINVENT souhaite INVENTorier, rendre visibles nationalement et internationalement et valoriser les importantes collections produites par les DISSidences « de l'Est » en exil en France, entre 1945 et 1991. Il s'agit dans le même temps de développer une réflexion sur la particularité de ces corpus, leur histoire et leur description en différentes langues. Les objectifs sont organisés en 2 axes :

1. identification des collections publiques et privées situées en France, créés par les dissidents et exilés venus d'URSS et des pays d'Europe centrale et de l'Est socialistes ; amélioration du signalement et de l'indexation ; échange d'expérience avec les institutions documentaires étrangères ; dissémination durable.
2. réflexion interdisciplinaire sur l'histoire comparée des circulations de collections diasporiques, la notion de patrimoine documentaire clandestin/militant/exilé ; conseils aux chercheurs sur le dépôt et les usages de ces corpus.

Dirigé par La Contemporaine et le laboratoire ICT Université de Paris, le projet inscrit les collections présentes en France dans une dynamique internationale de la recherche qui met l'accent sur le renforcement de la collaboration entre les chercheurs et les établissements patrimoniaux, travaillant en commun sur le document, numérique ou non, en ligne ou non. Il souhaite contribuer au futur réseau national de cartographie CollEx par des notices et l'identification de collections labellissables. L'objectif est aussi de renforcer la présence française dans les projets européens visant à inventorier et numériser la mémoire de la dissidence.

English summary : The DISSINVENT project aims to INVENTorate, make visible nationally and internationally and enhance the important collections produced by the Eastern DISSidents exiled in France between 1945 and 1991. At the same time, it involves a reflection on the particularity of this corpus, its history and its description in different languages. The objectives are organized into 2 axes :

1. identification of the public and private collections now located in France, produced by dissidents and exiled coming from USSR and from Eastern and Central European former socialist countries ; improvement of reporting and indexing ; exchange of experience with foreign documentary institutions ; sustainable dissemination.
2. interdisciplinary reflection on comparative history of the circulation of diasporic collections and the concept of clandestine / activist / exiled documentary heritage ; advice to researchers about the deposit and the uses of this corpus. Directed by La Contemporaine and the ICT laboratory at the University of Paris, the project places the collections present in France in an international research dynamic which emphasizes the strengthening of collaboration between researchers and major heritage institutions, which work together on the document, digital or not, both and online and offline. It wishes to contribute to the future national CollEx cartography network through notices and the identification of collections that could be labeled. The objective is also to strengthen the French collaboration to European projects aiming to inventory and digitalize the memory of dissent.

ÈS LETTRES

Université Paris Panthéon-Sorbonne

François Haas (Bibliothèque interuniversitaire de la Sorbonne – UP1) et **Pierre Verschueren** (Université de Franche-Comté)

ÈS LETTRES est un projet d'étude et de valorisation des thèses en lettres soutenues en France au XIXe siècle. Il vise à rassembler des informations relatives à ces documents (actuellement dispersées entre de multiples sources), procéder à la numérisation de ces thèses et préparer celle de documents qui leur sont associés (notamment les rapports de soutenance), élaborer une bibliographie à la fois générale et spécialisée sur ces thèses, constituer une base de données en ligne et reliée à des référentiels à partir de l'ensemble de ces éléments, et valoriser scientifiquement ces matériaux.

English summary : ES LETTRES is an enhancement and research project about doctoral dissertations defended in France during the 19th century in the fields of humanities and social sciences. It aims to bring together dispersed information about these documents, prepare digitization of theses and associated materials (such as the jury reports), make general and specialized bibliographies, create an online database using these elements, link it to datasets and enhance its scientific use.

Familielettres

Conservatoire national des arts et métiers

Cécile Formaglio (CNAM) et **Michel Lallement** (CNAM)

Ce projet baptisé Familielettres se donne pour objectif de créer un site sur la plateforme EMAN pour donner accès à la correspondance numérisée et en partie transcrite de Jean-Baptiste André Godin (1817-1888), industriel et créateur de l'expérimentation sociale inspirée du fouriérisme que fut le Familistère de Guise et de son épouse Marie Moret (1840-1908). Cet ensemble, conservé à la fois au Cnam et au Familistère de Guise, est particulièrement riche par les thèmes qu'il aborde mais aussi par sa complétude puisqu'il comprend aussi bien la correspondance active que passive. Le projet est porté par la Direction des bibliothèques et de la documentation du Cnam et le Familistère, avec deux laboratoires du Cnam, le Lise (Laboratoire interdisciplinaire pour la sociologie économique, UMR3320) et le Cedric (Centre d'études et de recherche en informatique et communications, EA4629), avec le partenariat de la plate-forme EMAN (Edition de manuscrits et d'archives numériques) développée au sein du laboratoire ITEM (UMR8132, CNRS Ecole normale supérieure). Le site qui en résultera sera aussi un terrain d'expérimentation sur les méthodes de visualisation d'un tel corpus et aboutira au développement d'une extension pour EMAN et la communauté Omeka.

English summary : The « Familielettres » project's primary goal is to create a site on the EMAN platform which will provide access to the digitized and partially transcribed correspondence of industrialist and Fourierist social experimenter who created the Familistère de Guise, Jean-Baptiste André Godin (1817-1888) and his wife Marie Moret (1840-1904). This collection, maintained by the CNAM and the Familistère de Guise, is particularly rich thanks to its thematic breadth and its completeness because it includes both sent and received correspondence. The project's participating organizations include the Direction des Bibliothèques et de la Documentation of the CNAM and the Familistère, two CNAM research centers (LISE and CEDRIC), and the EMAN platform developed by the ITEM laboratory (UMR8132, CNRS Ecole Normale Supérieure). The site will also serve as a testbed for developing methods to visualize such a large digital collection and will result in the development of an extension for EMAN and the Omeka community.

Féminismes en revue

Université d'Angers

Nathalie Clot (BU d'Angers), **Magali Guaresi** (Université Côte d'Azur) et **Christine Bard** (Université d'Angers)

Le projet « Féminismes en revue » vise à mettre sur pied une plateforme de recherche collaborative pour la numérisation, la structuration, l'enrichissement de métadonnées, l'annotation sémantique, la conservation et la diffusion de corpus de seize périodiques féministes du second XXème siècle (1944-2019).

Les objectifs poursuivis sont doubles, à la confluence d'une exigence patrimoniale et d'une demande scientifique. D'une façon générale, les archives de femmes et féministes souffrent de lacunes et de difficultés de conservation, d'archivage et de diffusion. Ce projet ambitionne de mettre à la disposition d'un large public des ressources périodiques concernant les mouvements féministes, adossées à des outils de traitement documentaire adaptés aux singularités de ces sources. Une attention sera portée aux unités de segmentation pertinentes, au traitement des illustrations ou encore au travail sur les métadonnées autrices/auteurs.

Ce programme vise également à nourrir les travaux scientifiques en études féministes en France et en Europe et à permettre le développement des recherches sur les revues féministes du XXème siècle en tant que dispositif d'expression et d'action privilégié. Dans une perspective de sociologie du travail féministe, ce corpus alimentera les études sur les contextes et les modes de production des périodiques (les institutions, l'organisation des tâches, le degré de professionnalisation des savoir-faire, le financement, etc.). Un axe de recherche s'organisera également autour de la question de la diffusion de ces objets, de leur réception et de leurs usages spécifiques par les lectrices. Le projet a enfin vocation à soutenir les études sur les courants féministes tant du point de vue de l'histoire politique et sociale que de l'histoire des idées.

English summary : The « Feminisms in Review » project aims to set up a collaborative research platform for the digitization, structuring, metadata enrichment, semantic annotation, preservation and dissemination of a corpus of 16 feminist periodicals of the second 20th century (1944-2019).

This project aims to two different goals : an heritage requirement and a scientific demand. Generally speaking, women's and feminist archives suffer from gaps and difficulties in conservation, archiving and dissemination. This project aims to make available, to a wide public, periodicals concerning feminist movements, backed by documentary processing tools adapted to the singularities of these sources. We will pay attention to the relevant segmentation units, to the processing of illustrations and to the work on author/author metadata.

This programme also aims to contribute to scientific work in feminist studies in France and in Europe and to allow the development of research on 20th century feminist journals as a privileged means of expression and action. From the perspective of the sociology of feminist work, this corpus will feed into studies on the contexts and modes of production of periodicals (institutions, organization of tasks, degree of professionalization of know-how, financing, etc.). A line of research will also be organised around the question of the dissemination of these objects, their reception and their specific uses by women readers. Finally, the project aims to support studies on feminist currents from the point of view of political and social history as well as the history of ideas.

Indexation Sémantique d'une archive scientifique et Services Associés pour la science ouverte

Centre de coopération internationale en recherche agronomique pour le développement
Anne Toulet (DIST Cirad) et **Franck Michel** (Université Côte d'Azur – CNRS - Inria)

Les services d'information scientifique et technique ont toujours été confrontés aux questions de gestion documentaire. Leur mission leur confère une responsabilité dans la conservation, la mise à disposition et la valorisation de ressources culturelles ou issues de la recherche. Ce projet s'intéresse à l'indexation sémantique des publications dans une archive ouverte et à l'exploitation de cette indexation pour développer des services aux chercheurs et aux documentalistes dans une perspective de science ouverte. La méthodologie mise en œuvre, adossée aux principes FAIR, se veut résolument générique : elle sera exploitable par toute communauté adoptant ces mêmes principes au travers de métadonnées interopérables, en particulier dans le cadre de systèmes de gestion de ressources documentaires. L'objectif du projet est double :

1. permettre une indexation automatique des publications avec des mots-clés thématiques et géographiques issus des référentiels standards (au format du Web sémantique) adaptés à chaque domaine et communauté scientifique ;

2. démontrer l'intérêt d'une telle démarche en développant des services innovants de recherche et de visualisation à destination des utilisateurs, capables d'exploiter cette indexation sémantique.

Agritrop, l'archive ouverte des publications du Cirad (<http://agritrop.cirad.fr>), servira de cas d'usage et de preuve de concept tout au long du projet. Les référentiels terminologiques utilisés ici seront en priorité le thésaurus Agrovoc, la base de connaissance collaborative Wikidata et le référentiel géographique GeoNames

English summary : Scientific and technical information services are concerned with the management of documentary resources. Specifically, they are responsible for the conservation, sharing and promotion of cultural resources produced by researchers. This project aims to improve access to and interoperability of the resources made available by IST services, while offering innovative services meant for documentalists and researchers in a multidisciplinary and open science mindset. The methodology, aligned with the FAIR principles, focuses on provisioning a generic solution usable by any community adopting these principles. This is achieved by means of interoperable metadata exploitable within documentary resources management systems. In this context, the goals of this project are twofold :

1. allow automatic indexing of documentary resources with thematic and geographic keywords from terminological resources (in the Semantic Web format) suitable for each domain or community ;

2. demonstrate the interest of this approach by developing innovative search and visualization services intended for users, capable of exploiting this semantic indexing.

Agritrop, Cirad's open publications archive (<http://agritrop.cirad.fr>), will serve as a use case and proof of concept throughout the project. The terminology resources will primarily be the Agrovoc thesaurus, the Wikidata collaborative knowledge base and the gazetteer GeoNames.

La guerre du droit

Université Paris 1 Panthéon-Sorbonne

Alexandra Gottely (Bibliothèque interuniversitaire Cujas – Paris 1) et **Florent Garnier** (Université Toulouse 1 Capitole)

Le présent projet est la deuxième étape d'une réalisation intitulée *Des facultés sur le front du droit* menée depuis 2017 grâce à une collaboration étroite entre le Centre toulousain d'histoire du droit et des idées politiques, le Centre de recherches critiques sur le droit de Saint-Etienne et la bibliothèque interuniversitaire Cujas. Le projet initial entendait saisir et comprendre la Grande Guerre à travers des juristes, leurs actions, leurs discours et leurs idées au sein d'institutions – facultés et universités – considérées séparément ou comparativement. L'objectif était à la fois de susciter des travaux de recherche inédits et de repérer et valoriser des ressources documentaires, iconographiques et archivistiques ; et de rendre l'ensemble, structuré, librement accessible sur internet. Dès l'origine, le projet se voulait évolutif et comprenait un appel à contributions collaboratif. Après avoir réussi à susciter et rassembler des recherches sur les juristes et les facultés de droit dans la Grande Guerre à Paris et Toulouse, puis Lyon et Bordeaux, l'objectif du présent projet est :

- la traduction de l'ensemble du site en anglais et en allemand pour soutenir et amplifier l'ouverture internationale,
- l'ajout de nouveaux partenaires, tant au niveau français, avec Lille, qu'europpéen, avec l'insertion de quatre facultés belges,
- le repérage et l'exploitation de nouveaux fonds documentaires, iconographiques et archivistiques,
- l'organisation d'une journée d'études pour valoriser les résultats obtenus.

English summary : This project is the second stage of an achievement entitled *Des facultés sur le front du droit* (*Faculties on the Law Front*), carried out since 2017 thanks to close collaboration between the Toulouse Center for the History of Law and Political Ideas (CTHDIP), the Center for Critical Research on the Law (CERCRID) of Saint-Etienne University and Cujas library. The initial project aimed to capture and understand the Great War through jurists, their actions, their speeches and their ideas within institutions - faculties and universities - considered separately or comparatively. The objective was both to stimulate new research work and to identify and enhance documentary, iconographic and archival resources ; and to make the whole, structured, reely accessible on the internet. From the outset, the project was meant to be progressive and included a collaborative call for papers. After having succeeded in generating and gathering research on jurists and law faculties in the Great War in Paris and Toulouse, then in Lyon and Bordeaux, the objective of today's project is :

- the addition of new partners, both at French level, with Lille, and European, with the integration of four Belgian faculties,
- identifying and using new documentary, iconographic and archival collections,
- the translation of the entire website into English and German,
- the organization of a study day.

METADICTMEDICA

Metadictionnaire médical multilingue de la bibliothèque numérique Medica

Université de Paris

Jean-François Vincent (BIU Santé – Université de Paris) et **Nathalie Rousseau** (CNRS – Sorbonne Université)

Au sein de la bibliothèque numérique Medica (BIU Santé), 49 dictionnaires médicaux et encyclopédies (XVIIIe-XXes.) sont interrogeables grâce à une interface spécifique : entrer un (début de) mot-clé fait accéder aux vedettes des articles, indexées dans les 331.643 pages des 453 volumes. Le projet, en produisant un nouvel outil et de nouvelles données, valorisera doublement ce trésor lexical et scientifique. L'utilisateur naviguera dans un corpus élargi à plusieurs dictionnaires majeurs du XVIIe s. (vedettes en grec et/ou latin) grâce à une nouvelle application qui donnera accès à la liste alphabétique continue de tous les mots du corpus (« métadictionnaire »). Les résultats d'une même requête seront considérablement enrichis : s'appuyant sur les données lexicographiques fournies par une dizaine de dictionnaires représentatifs, dans les corps d'article (variantes orthographiques, étymologies, traductions, synonymes, antonymes, hyperonymes et hyponymes, renvois divers) comme à la fin des volumes (index et glossaires), l'application suggérera, à travers une interface intuitive, des liens entre des mots actuellement dissociés par l'orthographe, par la langue (français, anglais, latin et grec ancien, mais aussi allemand, espagnol, italien), ou encore par l'évolution des usages. Par ailleurs, les données lexicographiques recueillies, matériau inédit d'un grand intérêt pour la recherche en histoire de la santé et pour l'étude de l'évolution des lexiques spécialisés en français ou du devenir des termes anciens dans nos langues scientifiques modernes, seront déposées en accès libre sur Medica.

English summary : Within the Medica digital library (BIU Santé), 49 medical dictionaries and encyclopaedias (18th-20thc.) are searchable through a dedicated interface : when entering a (beginning of a) keyword, the user is given access to the headwords of the entries which have been indexed in the 331,643 pages of the 453 volumes. By producing a new tool and new data, the project will enhance the value of this lexical and scientific treasure in two ways. Users will browse a corpus expanded to several major 17th-c.dictionaries (with Greek and/or Latin headwords) through a new application that will give access to a rolling alphabetical list (word wheel) containing all the headwords of this corpus ("metadictionary"). Furthermore, the results of a single query will be significantly enriched : on the basis of the lexicographical data provided by a set of representative dictionaries, in the entries (spelling variants, etymologies, translations, synonyms, antonyms, hyperonyms and hyponyms, various cross-references) and at the end of some volumes (index and glossaries), the application will suggest, through an intuitive interface, links between words currently dissociated by spelling, by language (French, English, Latin and Ancient Greek, but also German, Spanish, Italian) or by the evolution of usage. In addition, the collected lexicographical data, a new material of great interest for research in health history as well as for the study of the evolution of specialized lexicons in French or the evolution of ancient terms in our modern scientific languages, will be repositied in free access on Medica.

NumBois

Centre de coopération internationale en recherche agronomique pour le développement
Sylvie Vago (DIST Cirad) et **Jean Gérard** (Unité de recherche BioWooEB du Cirad)

Le Cirad dispose d'une xylothèque, une des plus importantes collections de bois tropicaux reconnue au niveau international avec plus de 34.000 échantillons pour 8.400 espèces botaniques. La base de données sur les caractéristiques technologiques de bois tropicaux et son application Tropix 7 complètent la Xylothèque. Cet ensemble représente un patrimoine scientifique et culturel remarquable sur le végétal, et un outil de recherche, d'expertise et de formation reconnu des spécialistes. En alimentant des ouvrages de vulgarisation sur les bois tropicaux, il s'adresse également à un large public avisé.

Par ailleurs, le Cirad s'est doté d'une bibliothèque numérique en agronomie tropicale, NumBA, lancée en 2019 dans le cadre d'un partenariat Gallica Marque blanche avec la Bibliothèque nationale de France. Le Cirad souhaite développer la valorisation de son patrimoine scientifique en enrichissant NumBA avec des données, informations et documents textuels et iconographiques sur les bois.

Le projet NumBois porte sur la numérisation de 3.000 échantillons de bois, sur l'enrichissement en métadonnées et l'intégration dans NumBA de ces échantillons et de fiches techniques, le reformatage du logiciel Tropix et son redéploiement sur internet en interrelation avec NumBA. Ce projet mené en coopération avec des universitaires africains, vise à valoriser des connaissances et des données uniques sur les bois en les diffusant en accès libre sur internet, via NumBA et Gallica pour consultation, réutilisation et partage par les internautes, universitaires, spécialistes, étudiants ou simples curieux.

English summary : The Cirad Xylotheque is one of the most famous collection of tropical and temperate wood samples in the world with 35,400 specimens representing more than 8,400 botanical species. In addition to the Xylotheque with its function of wood identification, the database of technological characteristics of tropical wood and its software Tropix 7 identify the technological characteristics of the main forest species.

This set constitutes a remarkable botanical and scientific heritage and a tool for research, expertise and training for wood and forests specialists. It also aims a large audience with the derived didactic documentary products on tropical woods.

In 2019, Cirad launched NumBA, its digital library on tropical agriculture in partnership with the Bibliothèque nationale de France. The documents in NumBA are indexed in Gallica. With this project "NumBois", Cirad wishes to develop the valorization of its cultural and scientific heritage on the internet and enriches NumBA with data, informations, textual and iconographic documents concerning woods. The so-called project concerns the digitization of 3000 samples of wood, metadata enrichment and the integration of the samples and technical sheets, the reformatting of Tropix for the internet in relation with NumBA.

The goal of this project, led in cooperation with African academics, is to provide access for free to a scientific base of knowledge with unique data on woods, via NumBA and Gallica to be consulted, re-used and shared by the internet users, whether users just having a look, students or academics

Pensée Anarchiste et Libertaire en Brochures

Etablissement Public Campus Condorcet

Alicia Léon y Barella (GED Condorcet), **Barbara Bonazzi** (CHS Mondes Contemporains) et **Paul Boulland** (CHS Mondes contemporains)

Porté par le Grand équipement documentaire du Campus Condorcet et par le Centre d'histoire sociale des mondes contemporains (CHS - UMR8058), le projet PALiBr vise à mettre à la disposition de la communauté scientifique un corpus numérisé et ocrisé d'environ 1200 brochures anarchistes francophones publiées de 1880 à 1918 en France, mais également en Suisse ou en Belgique, terres d'accueil d'anarchistes en exil.

Ce corpus dispersé au sein de différents établissements partenaires du CODHOS sera consultable sur la plateforme dédiée aux sciences humaines et sociales, celle de la bibliothèque numérique du Campus Condorcet.

Au-delà de l'intérêt patrimonial, ce corpus virtuel offrira une vue exhaustive de la production anarchiste de la période permettant d'explorer de nouvelles pistes de recherche, en croisant les points de vue disciplinaires.

Les données produites seront exploitées à moyen terme, dans le cadre d'un projet d'édition numérique en TEI (CELLF et CHS) et à plus long terme dans un projet de cartographie des réseaux anarchistes en Europe (CHS et CODHOS). Le projet PALiBr sera mis à profit pour permettre la montée en compétences des partenaires concernés, en particulier autour de l'encodage en XML-TEI, des outils de cartographie, de data visualisation et de l'élaboration des plans de gestion de données.

English summary : This project led by Campus Condorcet's Grand Equipement Documentaire (GED) and by the Centre d'Histoire Sociale des mondes contemporains (CHS – UMT8058) aims to digitize 1200 anarchist pamphlets to be electronically edited, searched, stored, displayed and made available to the scientific community. The corpus is composed of anarchist pamphlets written in French and published between 1880 and 1918 in France, but also in Swiss and Belgium, both welcoming nations for exiled anarchists.

The project aims at collecting texts preserved by different institutions members of the Collectif des centres de documentation en histoire ouvrière et sociale (CODHOS) in order to create a searchable virtual corpus hosted by Campus Condorcet's digital Library, a platform conceived for social sciences and humanities.

Beyond the cultural heritage management aspect, the virtual corpus will provide an overview of the anarchist production of the period and will allow to explore new research themes and practices through different approaches.

In the medium term, data will be proceeded for a digital TEI encoding edition (Centre d'étude de la langue et des littératures françaises CELLF and CHS), while in the long term, data concerning the anarchist network in Europe will be gathered, managed and analysed through a geographic information system (CHS and CODHOS). PALiBr project will contribute to the rise in skills and training of the partners especially as the XML-TEI encoding, GIS tools, data visualisation and Data Management Plan are concerned of a study day.

Projet d'Océrisation des recensements de la Population Parisienne

Centre national de la recherche scientifique

François Merveille (GED Condorcet) et **Sandra Brée** (LAHRA – UMR5190)

Ce projet vise à élaborer une vaste base de données (12 millions d'individus) à partir des recensements nominatifs de Paris de 1926, 1931, 1936 et 1946 qui sont les seuls recensements de la population parisienne existant avant la fin du 20^e siècle. Ces recensements sont déjà numérisés et en ligne sur le site des archives de Paris mais il s'agit ici de créer une base de données permettant l'exploitation statistique de ces documents particulièrement riches. Ce projet a deux intérêts principaux : méthodologique et scientifique.

Au niveau de l'enrichissement de la connaissance scientifique, l'analyse de ces recensements permettrait de connaître la population parisienne pendant l'Entre-deux-Guerres dans son intégralité. L'analyse quantitative et statistique des populations est rendue particulièrement difficile en histoire en raison du manque de bases de données et l'indexation des recensements de Paris de 1926, 1931, 1941 et 1946 offrirait l'opportunité d'un pas important dans la connaissance de la population urbaine européenne jusqu'alors très peu étudiée.

Mais surtout ce projet ouvrirait la voie à de nombreuses autres recherches grâce à l'utilisation et l'amélioration de nouvelles techniques de reconnaissance optique de caractères. Cette technique de lecture optique des manuscrits est une méthode novatrice, en amélioration constante. Le traitement que l'on pourra appliquer aux données manuscrites des recensements de population parisiens pourra ensuite être dupliqué sur de nombreux documents, permettant ainsi de mettre à disposition des milliers de données.

English summary : This project aims to develop a vast database (12 million individuals) based on the Paris nominal censuses of 1926, 1931, 1936 and 1946, which are the only censuses of the Parisian population existing before the end of the 20th century. These censuses have already been digitized and put online on the Paris Archives website, but the aim here is to create a database for the statistical exploitation of these particularly rich documents. This project has two main interests : methodological and scientific. At the level of enriching scientific knowledge, the analysis of these censuses would make it possible to know the Parisian population during the Interwar period in its entirety. The quantitative and statistical analysis of populations is made particularly difficult in history because of the lack of databases. The indexing of the 1926, 1931, 1941 and 1946 Paris censuses would offer the opportunity of a major step forward in the knowledge of the European urban population, which had been little studied until then.

Above all, this project would pave the way for much further research through the use and improvement of new OCR techniques. Optical Character Recognition is an innovative method that is constantly being improved. The treatment that can be applied to the handwritten data of the Paris population census can then be duplicated on many documents, thus making thousands of data available.

Transcrire

Musée du Quai Branly – Jacques Chirac

Paul-Emmanuel Bernard (MQB-JC) et **Anaïs Wion** (Institut des mondes africains – UMR8171)

La plate-forme *Transcrire* est un outil de transcription collaborative. Elle propose à chacun (chercheurs comme publics amateurs) de contribuer à la transcription de documents manuscrits déjà numérisés en mode image et provenant de collections patrimoniales de bibliothèques de recherche. Son périmètre éditorial couvre les matériaux de terrains manuscrits provenant de fonds d'archives de chercheurs en sciences humaines et sociales (carnets de terrain, carnets de fouille, notes de voyage et d'enquête par exemple).

Le projet vise à refondre, stabiliser et étoffer l'offre de service liée à la transcription de ces corpus manuscrits. La migration du site web vers un environnement Omeka S doit permettre à la fois d'anticiper l'obsolescence technologique de la plateforme actuelle et d'ancrer celle-ci dans l'ouverture des données et le web sémantique.

Il s'agit aussi, par l'ajout de fonctionnalités nouvelles, la mise en place d'actions de formation et la production de livrets d'accompagnement, de donner une plus grande autonomie aux bibliothèques et projets de recherche dans :

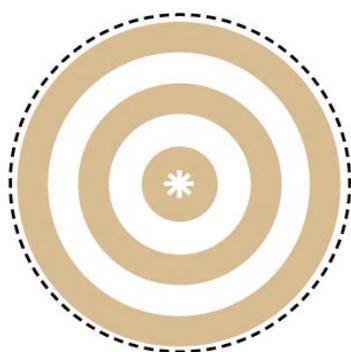
- la gestion des collections, des métadonnées et des transcriptions ;
- la mise en place de chaînes d'intégration des transcriptions dans leurs systèmes d'information ;
- l'articulation de la plate-forme avec des solutions de reconnaissance automatique d'écriture manuscrite ;
- l'animation de communautés de contributeurs.

English summary : The *Transcrire* platform is a collaborative online transcription tool. It allows researchers as well as amateurs to contribute to the transcription of already digitized manuscript documents held by research libraries in their heritage collections. It gathers manuscripts produced by researchers in humanities and social sciences during their fields : for instance field notebooks, excavation notebooks, travel and survey notes.

The project aims at redesigning, stabilizing and expanding the services related to the transcription of these manuscript corpus. The migration of the website to an Omeka S environment should allow both to anticipate the technological obsolescence of the current platform in Omeka and to anchor it in the opendata and the semantic web.

Furthermore, the addition of new functionalities, the development of training actions and the production of guidelines and booklets, will empower libraries and research projects

- in their management of collections, metadata and transcripts ;
- in the creation of processing chains in order to integrate the transcripts into their own information systems ;
- in the articulation of the platform with automatic handwriting recognition solutions ;
- in facilitating the emergence of communities of contributors.



En savoir plus sur l'AAP19_20 :
collexpersee.eu , rubrique « **Projets** ».

Pour suivre notre actualité,
Rendez-vous sur Twitter
@COLLEX_IR