Ar(abes)ques
JANVIER - FÉVRIER - MARS 2022

DOSSIER

CollEx-Persée
Un dispositif au service de l'excellence documentaire

104

PLEINS FEUX SUR • Le GED Condorcet

SYSTÈME D • Koha, un système qui fait ses preuves

abes
agence bibliographique de l'enseignement supérieur

(Dossier) CollEx-Persée : un dispositif au service de l'excellence documentaire

« Imaginé à partir de 2014 pour remplacer le système des Cadist (Centres d'acquisition et de diffusion de l'information scientifique et technique) devenu obsolète, le groupement d'intérêt scientifique (GIS) CollEx-Persée, créé officiellement en 2017, a profondément bouleversé le paysage de l'information scientifique et technique française. Alors que le GIS a publié en décembre 2021 son autoévaluation et que bat son plein le travail de réflexion pour donner naissance, à partir de 2023, à CollEx-Persée 2, le dossier de ce numéro d'*Arabesques* intitulé « CollEx-Persée : un dispositif au service de l'excellence documentaire » propose de faire le point sur les enjeux, les problématiques et les perspectives de cet outil singulier, et présente plusieurs projets documentaires réalisés dans le cadre qu'il a permis d'installer.
Ce premier numéro de l'année 2022 inaugure également une nouvelle rubrique, « Système D, un outil pour vos données », suite aux suggestions faites par nos lecteurs dans l'enquête de lectorat menée en 2021. Bonne lecture ! Bonne année 2022 ! »

04 CollEx-Persée : passé – présent – avenir Alain Colas

06 CollEx-Persée, retour sur une dynamique de réussite
Catherine Desos-Warnier

08 Les grandes orientations pour CollEx-Persée 2 Alain Colas

09 Interview de David Aymonin et Marianne Giloux
« CollEx-Persée a bouleversé le paysage de l'IST française »

10 Persée : numériser pour la recherche Gabrielle Richard

12 Cartographie CollEx-Persée : valoriser les « Collections d'Excellence » Clément Oury et Philippe Goguely

14 Datacc : une approche renouvelée de la notion de données, de et pour la Recherche
Amanda Cruguel-Vitry et Marie-Emilia Herbet

15 NumBois, rencontre entre la recherche en sciences du bois et l'information scientifique et technique
Sylvie Vago et Marie-Françoise Fily

16 FamiliLettres : un accès unique à un corpus de correspondances exceptionnelles
Cécile Formaglio

17 La République locale des lettres : analyse du jeu littéraire et militant en Bretagne Mannaig Thomas et Philippe Lagadec

18 Mistara : au croisement des enjeux de la transition bibliographique et de la recherche sur l'onomastique - ou étude des noms - arabo-musulmane Cécile Gobbo

19 Datapoc 2.0 : un référentiel « Personnes » spécifique au Muséum national d'Histoire naturelle
Chloé Besombes, Cécile Callou et Cindy Lim

20 Les FID allemands : vers une infrastructure nationale d'information Marianne Dörr

22 La conservation partagée des périodiques de l'enseignement supérieur : améliorer l'accès à la documentation Guillaume Niziers

23 Les archives du XIXe siècle du Parlement britannique dans Calames Étienne Naddeo et Yann Nicolas

24 (Pleins feux sur...)
Le GED Condorcet : un objet documentaire unique dédié aux sciences humaines et sociales
Stéphanie Groudiev

26 (Système D...)
des outils pour vos données
Koha, un système qui fait ses preuves
Sonia Bouis

27 (Actus...)
• Des données plus riches et mieux structurées : en janvier 2022, le format Unimarc/B évolue
Laure Jestaz

• « Booksprint » et catalogage des ressources continues en ligne au menu de la journée des CR
Yves Desrichard

28 (Portrait)

Ar(abes)ques
REVUE TRIMESTRIELLE DE L'AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPÉRIEUR,
227, avenue du Professeur-Jean-Louis-Viala, CS 84308, 34193 Montpellier cedex 5.
Tél. 04 67 54 84 10 / Fax 04 67 54 84 14 / https://abes.fr
Directeur de la publication : David Aymonin.
Coordination éditoriale et secrétariat de rédaction : Véronique Heurtematte.
Comité de rédaction : Christophe Arnaud, Aurélie Faivre, Christine Fleury, Morgane Parra Raluca Pierrot, Laurent Piquemal, Marie-Pierre Roux.
Iconographie rassemblée par Christophe Arnaud.
Conception graphique : Anne Ladevie (anneladevie.com). Impression : Pure Impression

Revue publiée sous licence Creative Commons CC BY-ND 2.0
(Paternité - Pas de modifications) sauf pour les images qui peuvent êtres soumises à des licences différentes ou à des copyrights.

Couverture : Loupe ancienne sur un vieux manuscrit, crédit photo scis65 - 123RF

Les opinions exprimées dans* Arabesques *n'engagent que la responsabilité de leurs auteurs.

ISSN (papier) 1269-0589 / ISSN (web) 2108-7016

En 2022, au revoir CollEx-Persée 1, bonjour CollEx-Persée 2

Le lancement du GIS CollEx-Persée fin 2017[1] a marqué à la fois la fin de la Bibliothèque Scientifique Numérique (BSN)[2] et la transformation en profondeur du réseau des Cadist[3], centres d'acquisition et de diffusion de l'information scientifique et technique, mis en place en 1983.

En une seule opération, les initiateurs de CollEx-Persée ont changé le paysage de l'IST française et apporté une innovation radicale en considérant les collections numériques et les collections papier des plus grandes bibliothèques universitaires françaises non plus comme des parties mais comme un tout constituant LA Collection d'Excellence, au service de la communauté nationale des chercheurs. Dès lors, bibliothécaires et chercheurs peuvent travailler ensemble cette matière pour l'enrichir, la structurer, l'explorer, créer d'autres données et connaissances.

Enfin une transformation radicale est à l'œuvre : avec CollEx-Persée, le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation confie la gestion de ses crédits et de sa politique à ceux qui en sont les bénéficiaires. L'argent que la Mission de l'information scientifique et technique et du réseau documentaire (MISTRD) – environ 5 millions d'euros par an – donnait aux Cadist est maintenant placé entre les mains des bibliothèques elles-mêmes, rassemblées au sein d'une gouvernance à la fois scientifique et technique multipolaire. Et ça marche !

Merci aux auteurs qui ont contribué à cette nouvelle et excellente livraison d'*Arabesques,* qui nous montrent comment CollEX-Persée 1 (2017-2022) s'est installé et a profondément modifié l'activité de nombreuses bibliothèques et laboratoires, et a déjà su créer nombre de nouveaux réservoirs de données, de nouveaux services, de nouveaux terrains d'exploration.

Merci également pour la sincérité et la transparence des propos d'Alain Colas, qui en tant qu'administrateur de la Bibliothèque nationale et universitaire et directeur du GIS CollEx-Persée sait et nous dit quelles montagnes sont encore à déplacer pour atteindre au but.

Tout d'abord, faire que les bibliothèques d'excellence se perçoivent comme les animatrices de réseaux

thématiques, en reliant entre elles les bibliothèques de leurs disciplines pour construire ensemble des services aux chercheurs et partager leurs collections.

Ensuite, faire en sorte que la numérisation et le signalement des collections soient réalisés à l'échelle qui convient.

Et que tout cela forme ce que j'ai l'immodestie d'appeler la « collection nationale », d'excellence parce qu'exhaustive, accessible, ouverte et partagée, utilisable parce que signalée et structurée, hybride entre papier, numérique et numérisée, dynamique car gérée avec les chercheurs et les citoyens pour répondre à leurs besoins.

C'est ce que dessine le projet CollEx-Persée 2 qui émergera en 2022. Formons le vœu qu'il se réalise avec l'aide de toutes ses parties prenantes et nous apporte encore tant d'innovations.

Bonne année 2022 à tou-te-s les professionnel-le-s de la documentation et de l'information,

Bonne année à tous nos publics.

David Aymonin

Directeur de l'Abes

[1] Comment ne pas signaler l'article d'Archimag et la savoureuse caricature qu'il contient, publiés en 2018 **https://www.archimag.com/bibliotheque-edition/2018/08/24/bibliotheques-universitaires-adieu-cadist-bonjour-collex**

[2] **https://publications-prairial.fr/arabesques/index.php?id=1357**

[3] **https://publications-prairial.fr/arabesques/index.php?id=940**

CollEx-Persée : *passé – présent – avenir*

Initié à partir de 2014 en remplacement des Cadist, CollEx-Persée amorce un tournant majeur avec un nouveau projet articulé autour de programmes à forte mutualisation.

Il y a un peu plus de 10 ans, au tournant des années 2010, la Mission de l'information scientifique et technique et du réseau documentaire (la MISTRD) au sein du ministère de l'Enseignement supérieur et de la recherche, avait amorcé une prudente révision des Cadist (Centres d'acquisition et de diffusion de l'information scientifique et technique), moins sur le fond touchant aux missions, que sur la forme. En effet, dans le prolongement d'une globalisation des moyens déjà à l'œuvre pour les établissements d'enseignement supérieur, il avait été décidé de soumettre une part de leurs dotations à des critères de « performance », pouvant aller jusqu'à 30 % du montant total. Beaucoup d'énergie avait été dépensée pour élaborer un système de notation devant permettre d'objectiver cette part sur critères, et qui s'avèrera être une vraie usine à gaz (je le dis sans ambages car j'étais moi-même à la manœuvre). Ainsi trouvait-on parmi les éléments d'évaluation : l'activité du Prêt entre bibliothèques (PEB), le nombre de jours d'ouverture par an, la participation ou non à un plan de conservation partagée, à des programmes de numérisation ou à un projet de recherche, la visibilité sur le Web en tant que Cadist, avec en regard, une échelle de notes…

Cette révision ne promettait pas les grands soirs, mais elle reflétait un questionnement latent au regard du maintien d'une dotation spécifique et néanmoins récurrente depuis le début des années… 1980. Je ne reviens pas sur les raisons de ce soutien qui, d'une certaine manière, venait palier la grande misère des bibliothèques universitaires en termes de moyens à l'heure de la massification des universités au tournant de ces années 1980. Certains iront jusqu'à parler d'un « cache-misère ».

Avec l'introduction de cette part sur critères, pas de révolution donc, à peine un ajustement des missions : fin de la notion d'exhaustivité pour les acquisitions dans un domaine donné, en particulier en langues étrangères, aucun des Cadist n'ayant les moyens d'assurer cette mission qui était centrale, et incitation à aller davantage vers les ressources électroniques. Avec un constat sévère : la lente et irrésistible baisse de l'activité du PEB, à quelques exceptions près, qui constituait l'autre mission fondatrice de ces centres. Réinterroger le modèle des Cadist s'imposait pourtant dans un environnement qui avait fortement évolué depuis leur création : la place prépondérante des ressources électroniques dans l'écosystème des bibliothèques universitaires, l'émergence de la question de l'*open access*, l'évolution des pratiques des chercheurs, l'autonomie des établissements.

UN CHANTIER ENTAMÉ EN 2014

De fait, c'est quelques années plus tard, en 2014, qu'a été engagé le véritable chantier de refonte des Cadist, suivant en cela, et avec quelque retard, la volonté du ministère de restructurer les politiques publiques en matière d'information scientifique et technique. Ce fut la création de la Bibliothèque scientifique numérique (BSN) dès 2007, inscrite, à l'initiative de l'Alliance Athéna, dans la feuille de route des infrastructures de recherche en 2012[1], aux côtés d'une nouvelle infrastructure dite « documentaire » : CollEx.

Cette démarche était fortement inspirée par le modèle des FID[2] allemands qui connurent une refonte importante à la même époque avec la priorisation des soutiens disciplinaires *via* des appels d'offre, une forte orientation vers le numérique (*ressources électroniques et numérisation*) et le développement des services à la recherche.

À partir de cette année 2014, une « comitologie » fournie s'est mise en place (groupe de travail, comité de préfiguration, comité opérationnel) pour aboutir à un nouveau dispositif qui se déploiera en deux temps :

- En 2016, à la faveur de l'actualisation de la feuille de route des infrastructures de recherche[3], la redéfinition de CollEx pour devenir CollEx-Persée, une infrastructure dite « en information scientifique et technique », englobant le potentiel des collections d'excellence et Persée.
- En 2017, au travers d'un double appel à manifestation d'intérêt (AMI) lancé par le ministère afin d'arrêter le périmètre de ces collections d'excellence : choix des membres du dispositif et labellisation de collections spécifiques.

UN PRINCIPE DE COCONSTRUCTION ENTRE BIBLIOTHÈQUES ET COMMUNAUTÉS SCIENTIFIQUES

Le dispositif qui en découlera sera novateur à bien des égards. D'abord la gouvernance et le pilotage

[3] https://www.enseignementsup-recherche.gouv.fr/sites/default/files/imported_files/documents/infras_def3_243296.pdf

[2] F*achinformationdienst* ou services d'information spécialisés, équivalent à une infrastructure d'information pour la recherche en Allemagne. https://publications-prairial.fr/arabesques/index.php?id=585

[3] https://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/74/5/feuille_route_infrastructures_recherche_2016_555745.pdf

sont déconcentrés au niveau des établissements et des opérateurs en IST par le biais d'un groupement d'intérêt scientifique (GIS) porté par la Bibliothèque nationale universitaire (Bnu) de Strasbourg. Le réseau se dote d'un conseil scientifique marquant la prééminence, notamment dans le cadre des projets soutenus par CollEx-Persée, du principe de coconstruction entre bibliothèques et communautés scientifiques.
Dans le même temps, il assume un continuum au travers du soutien aux acquisitions disciplinaires ou thématiques à une vingtaine de bibliothèques sur le modèle Cadist, ainsi qu'à des programmes d'intérêt national en grande partie déjà amorcés par BSN et se traduisant dans le cadre de CollEx-Persée par des actions très concrètes : numérisation, signalement, acquisitions de ressources électroniques en licence nationale, plans de conservation partagée, archivage pérenne.
Au terme de quatre années de fonctionnement, et au regard de son jeune âge, CollEx-Persée est un pari réussi. Celui-ci s'apprécie tout particulièrement au travers des programmes et projets collaboratifs dont se sont emparées les communautés scientifiques. En revanche, on ne peut que constater que les réseaux autour des « collections disciplinaires » ne se sont pas développés, les politiques d'acquisition sur les dotations récurrentes aux bibliothèques délégataires et associées ne dépassant pas le plus souvent le périmètre local.

BÂTIR UN PLAN DE NUMÉRISATION À LA HAUTEUR DES ENJEUX DE L'ESR

Les orientations qui émergent des travaux de l'autoévaluation[4] du GIS posent la vision d'un dispositif transformant pour les bibliothèques, basé davantage sur la donnée et les services. De fait, CollEx-Persée doit assumer pleinement sa fonction d'infrastructure en information scientifique en favorisant l'inscription des gisements documentaires dans les projets de recherche.
Au centre des enjeux, la crise sanitaire a montré l'absolu besoin d'une offre massive de ressources numériques accessibles à distance. Il devient urgent pour l'Enseignement supérieur et la recherche (ESR) de se doter d'un plan ambitieux de numérisation à hauteur des collections détenues par les bibliothèques afin de rattraper le retard pris par la France par rapport à des pays comparables dans le monde. Ce programme majeur, qui nécessitera des moyens conséquents, actera le rôle prépondérant que joue déjà Persée au sein du GIS.
Le nouveau dispositif devrait marquer, à compter de 2023, la transition vers un modèle mutualisé qui sera celui d'une véritable infrastructure répartie. Celle-ci, dans sa nouvelle organisation, sera appelée à être plus inclusive et à favoriser un effet réseau à grande échelle au travers de programmes particulièrement structurants et à forte mutualisation. Avec ce modèle ouvert, propice à l'innovation, il s'agira de sortir d'un périmètre figé de disciplines qui avait peu évolué depuis près de quarante ans sur la base des Cadist thématiques.
Sa réussite dépendra de la capacité du prochain GIS CollEx-Persée à répondre à plusieurs conditions. En premier lieu, ce dispositif, au service de tous les acteurs de l'ESR, devra être porté par une vraie volonté politique, en particulier au niveau des responsables des établissements. Ce qui suppose que les futurs programmes apportent une réponse tangible aux stratégies mêmes de ces établissements en matière de documentation et d'information scientifique, et rencontrent l'intérêt des infrastructures et opérateurs spécialisés. L'inscription des programmes dans l'environnement de la science ouverte sera, de fait, essentielle, en synergie avec tous les acteurs dans ce domaine.

Alain Colas
Directeur de CollEx-Persée
alain.colas@bnu.fr

[4] Le bilan à mi-étape du GIS CollEx-Persée a fait l'objet d'un véritable chantier d'autoévalution sur toute l'année 2020 et une partie de 2021, impliquant des entretiens avec les principaux acteurs de la science ouverte et de l'information scientifique, et en s'appuyant sur deux groupes de travail : **https://www.collexpersee.eu.**

CollEx-Persée, *retour sur une dynamique de réussite*

Conscient de l'enjeu de s'inscrire durablement dans le paysage de l'information scientifique et technique (IST), CollEx-Persée a entrepris son bilan. Cette autoévaluation doit permettre de poser, collectivement, les bases solides d'un nouveau projet.

Le dispositif CollEx-Persée entend faciliter l'accès des chercheurs aux collections d'excellence détenues par les grandes bibliothèques scientifiques et encore trop peu exploitées. Cette ambition, matérialisée dans un GIS (groupement d'intérêt scientifique) en 2017, a été reconnue par l'Enseignement supérieur, la Recherche et l'Innovation dans le cadre du récent renouvellement de son statut d'infrastructure de recherche, pour la période 2021-2025. Ce label représente une reconnaissance de l'importance nationale et stratégique de CollEx-Persée et l'incite à collaborer étroitement avec les autres infrastructures de recherche. Tout n'est cependant pas acquis et, conscient de l'enjeu de s'inscrire durablement dans le paysage de l'information scientifique et technique (IST), CollEx-Persée a entrepris son bilan. Cette autoévaluation[1] doit permettre de poser, collectivement, les bases solides d'un nouveau projet.

DES OBJECTIFS TOUJOURS D'ACTUALITÉ

L'infrastructure vise à faire des collections d'excellence, des matériaux scientifiques intégrés dans des projets de recherche, en les rendant bien plus visibles qu'elles ne le sont aujourd'hui, par leur labellisation, leur cartographie, leur signalement enrichi de haut niveau, leur valorisation et leur accès facilité. Il s'agit de ressources hybrides, dans la mesure où CollEx-Persée encourage la production et la collecte d'objets numérisés et nativement numériques qui viennent compléter, enrichir ou documenter toute collection sur support.

À partir de ces gisements, il s'agit de développer de nouveaux services en partenariat étroit avec la recherche, inscrits dans les principes de la science ouverte et bien environnés. Cette réflexion revient à positionner les professionnels comme des acteurs de la recherche à part entière et à valoriser leur expertise et leurs compétences dans l'exploitation et la gestion des données et des documents.

En conséquence, les questions liées à l'ensemble de la chaîne de traitement, depuis la diffusion ouverte jusqu'aux questions de préservation pérenne, font partie du champ d'action de CollEx-Persée, qui les instruit à partir de programmes fédérateurs, au service de l'ensemble de la communauté.

UNE MISE EN ŒUVRE COCONSTRUITE AVEC LES CHERCHEURS

Le défi a été de mobiliser un réseau réparti sur tout le territoire national, comprenant les grandes bibliothèques de recherche, les opérateurs en IST, la Bibliothèque nationale de France et les institutions dotées de collections d'excellence. Le souhait était de pouvoir offrir à tous ces acteurs un lieu de rencontres, de valorisation et de partage.

Pour se structurer, l'originalité est d'avoir construit

[1] https://www.collexpersee.eu/le-rapport-dautoevaluation-de-collex-persee-est-en-ligne

[2] Ces programmes sont évoqués par leurs pilotes dans d'autres parties de ce numéro d'Arabesques.

© Ludwig Uhland - Collection Bnu – Numistral

Manuscrits allemands. Dessins d'animaux, exécutés par le poète Ludwig Uhland, enfant.

© Collection Bnu

Manuscrit jaïna Nandicurni de Jinadasa, XIVe siècle.

une gouvernance mêlant les profils et compétences à travers trois organes : le conseil des membres, le conseil scientifique et le comité opérationnel. Les différents programmes[2] sont instruits par des groupes de travail dotés d'une feuille de route ajustée régulièrement avec les instances. Une nouvelle fonction a été créée en 2019, celle des référents CollEx au sein des établissements, pour relayer les actions du GIS. Le panel d'experts s'est, par ailleurs, étoffé au gré des sollicitations dans le cadre des divers travaux menés par le dispositif : la communauté compte ainsi une cinquantaine de professionnels mobilisables sur différentes thématiques. Enfin, dans le cadre de ses actions, CollEx-Persée est en relation avec de nombreux partenaires, s'enrichissant de ces échanges : l'ADBU, Couperin, Istex, l'Enssib, l'Urfist, Huma-Num, le Siaf, le Cines, le Rnmsh, HAL, etc. L'ensemble est piloté par la Bibliothèque nationale et universitaire, qui assure la gestion de la dotation annuelle du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation de 5 millions d'euros.

Une des réussites majeures de CollEx-Persée est la forte implication des chercheurs dans les orientations prises durant les quatre ans écoulés. Ils ont été très présents pour accompagner les actions du GIS, par exemple dans la sélection des licences nationales à négocier, ou pour faire émerger et soutenir des projets à vocation nationale comme ResPaDon[3] et Ouvrir CartoMundi[4]. Ils ont été essentiels dans la sélection de nouveaux projets de numérisation et de services à la recherche, via les appels à projets du dispositif, mécanisme qui a eu un effet transformant dans les pratiques. Trois appels à projets (le troisième est en cours) et un appel à résidence de chercheurs ont ainsi permis de mobiliser les acteurs de la recherche et de l'IST sur des actions communes, de valoriser des fonds d'excellence et de produire des résultats utiles à une vaste communauté. C'est une part importante dans le bilan financier du GIS puisque, en 5 ans, près de 5,9 millions d'euros auront été engagés pour cet axe.

INTERROGER LA DYNAMIQUE EN COURS

Il avait été convenu, à mi-parcours du GIS, d'établir un bilan pour construire un nouveau projet. Parmi les objectifs atteints, on constate une réelle montée en compétences des professionnels de l'information en ingénierie de projets, ainsi que l'image positive du réseau auprès des chercheurs pour lesquels il vient pallier un manque dans la conduite de leurs recherches[5], ou encore la mise en place de programmes nationaux transverses aux communautés et la capacité du dispositif à avoir un rôle structurant, de par la coordination qu'il assure entre tous les acteurs de l'information scientifique.

Les points d'attention relevés sont, quant à eux, de plusieurs ordres. Pour assurer une meilleure lisibilité du dispositif, il est préconisé de reformuler la stratégie globale du GIS et de mieux afficher l'engagement de la gouvernance. Le modèle financier est à rénover, car trop segmenté et, pour moitié, encore fléché sur les anciens Cadist (Centres d'acquisition et de diffusion de l'information scientifique et technique). Outre la pérennisation de l'enveloppe initiale, indispensable si l'on souhaite des infrastructures scientifiques de haut niveau que le seul mode projet ne peut permettre, il conviendra de trouver des sources de financement complémentaires pour de futurs programmes nécessitant un investissement massif.

Un double enjeu demeure encore : mieux connaître les usages des nouveaux services développés et favoriser la mise en réseau et la réplicabilité de ces mêmes services au plus près des besoins des chercheurs. Enfin, la nécessité de s'ouvrir vers l'extérieur en recherchant des collaborations au sein de différents réseaux européens a été soulignée. Le récent congrès qui s'est tenu à la BNU avec les collègues allemands des *Fachinformationdienste* a, à cet égard, rappelé que le travail scientifique se caractérisant par l'internationalisation, il serait judicieux que des projets transfrontaliers soient développés[6].

Suite à ces travaux d'autoévaluation, un rapport de synthèse a été soumis au conseil des membres du 15 octobre 2021, qui a validé les pistes de travail proposées en vue d'un nouveau projet.

UNE ANNÉE 2022 QUI S'ANNONCE INTENSE

La transition du dispositif est donc en cours et l'année à venir sera fondamentale pour sa mise en œuvre. Les orientations choisies tournent principalement autour de la notion de programmes à forte mutualisation, dotés d'une gouvernance propre, impliquant les opérateurs, permettant des économies d'échelle et un développement d'envergure des outils utilisés, des contenus produits et des réseaux constitués pour les porter. La définition de ces programmes aura, par ailleurs, des conséquences sur l'organisation générale et le modèle économique.

Il faut aboutir à un nouveau projet pour 2023, tout en prenant le temps de construire collectivement et d'accompagner ce changement avec le réseau. Le contexte est cependant propice car les établissements sont déjà en train d'investir de nouveaux champs d'actions basés sur la coopération renforcée entre bibliothécaires et chercheurs, enjeu du XXIe siècle. CollEx-Persée peut donc être vu comme un accélérateur, un facilitateur ou un soutien dans cette évolution stratégique de la place des bibliothèques dans les universités.

Catherine Desos-Warnier
Chargée de mission pour le GIS-CollEx-Persée
catherine.desos-warnier@bnu.fr

CHIFFRES CLÉS

- **5 Millions d'euros par an**
- **21 établissements, 4 opérateurs, la Bibliothèque nationale de France**
- **170 labels d'excellence**
- **Un réseau à 87 % en sciences humaines et sociale**
- **5 groupes de travail**
- **23 référents CollEx**
- **59 projets collaboratifs numérisation et services (dont 35 achevés)**
- **7 résidences de chercheurs (2021-2022)**
- **2 projets à dimension nationale (2021-2022)**
- **8 corpus en licence nationale acquis, 12 en négociation**
- **1 126 000 nouvelles vues numérisées (septembre 2021)**
- **31 projets déposés dans le dernier appel à projets**

[3] ResPaDon : https://www.collexpersee.eu/projet/respadon

[4] Ouvrir CartoMundi : https://www.collexpersee.eu/projet/ouvrir-cartomundi

[5] *Cf.* l'ensemble de témoignages de chercheurs recueillis dans la rubrique « #Paroledechercheurs » du GIS. A retrouver dans https://www.collexpersee.eu/rubriques

[6] Pour la restitution de ce congrès, *cf. BBF,* 30/09/2021. En ligne : https://bbf.enssib.fr/tour-d-horizon/rencontres-collex-persee-fid_70100

Les grandes orientations pour CollEx-Persée 2

En 2022 s'ouvrira un nouveau chantier, celui de la coconstruction de programmes avec les établissements qui se porteront volontaires dans le cadre d'un pré-appel à manifestation d'intérêt.

DES ORIENTATIONS FORTES RESSORTENT DES TRAVAUX DE L'AUTOÉVALUATION :

• CollEx-Persée s'organisera autour de programmes à forte mutualisation. Leur nombre n'excèdera pas une dizaine. Il s'agira de programmes pluriannuels dont la durée pourra aller jusqu'à 5 ans. Parmi ceux-ci, des programmes blancs feront l'objet d'un appel à manifestation d'intérêt spécifique et seront l'occasion de coconstructions avec les établissements/bibliothèques.

• Le nouveau dispositif induira une forte responsabilisation des acteurs du terrain, en termes de livrables, d'animation de réseaux, de valorisation des résultats mais aussi de capacité de pilotage, de gouvernance et de gestion budgétaire.

• Le plan de numérisation pour l'Enseignement supérieur et la Recherche constituera le programme phare et nécessitera un financement externe massif. Le rôle central de Persée est réaffirmé.

• Les missions de CollEx-Persée seront en partie recentrées sur les processus métiers à forte valeur ajoutée, en adéquation avec les pratiques des chercheurs et facilitant l'intégration des gisements documentaires dans les projets de recherche. Ils devront répondre à des critères partagés : un potentiel d'ouverture large ; des livrables reproductibles pour une dissémination facilitée ; la garantie de faire réseau, en particulier en s'appuyant sur le maillage des collections labellisées et en incluant les communautés scientifiques ; une hybridation des compétences privilégiée.

• Les opérateurs en information scientifique auront toute leur place dans le développement des programmes (Abes, CTLes, Inist-CNRS). Un partenariat renforcé avec des opérateurs de la culture sera recherché (BnF, Archives de France).

• CollEx-Persée s'inscrira d'une manière volontariste dans les principes de la science ouverte. Des synergies seront à formaliser avec les autres infrastructures et plateformes intervenant dans ces domaines : Huma-Num, Istex, entrepôt national des données.

• Une attention particulière sera portée au soutien et à la valorisation des disciplines rares au travers de projets documentaires spécifiques en appui aux communautés scientifiques concernées.

• CollEx-Persée a vocation à contribuer au développement de sujets transversaux : formation et montée en compétences, questions juridiques, diffusion de la culture scientifique.

• CollEx-Persée devra s'inscrire dans l'environnement international, en particulier européen où il peut d'ores et déjà s'appuyer sur les relations privilégiées avec les FID allemands. Il aura à se positionner par rapport aux universités européennes et à affirmer sa présence au sein des associations professionnelles internationales telles que Liber et l'IFLA.

Alain Colas
Directeur de CollEx-Persée
alain.colas@bnu.fr

© Jean-Pierre Rosenkranz

Photo de groupe lors de la journée des référents de novembre 2021

INTERVIEW

David Aymonin, directeur de l'Abes
et **Marianne Giloux,** responsable du département
Métadonnées et services aux réseaux de l'Abes

« COLLEX-PERSÉE TRANSFORME LE PAYSAGE DE L'IST FRANÇAISE »

***Arabesques* : Quels sont les grands axes de la collaboration entre CollEx-Persée et l'Abes ?**

DAVID AYMONIN : à sa création, CollEx-Persée a pu sembler un objet non identifié mais la vraie nouveauté de ce dispositif est apparue rapidement : relier les attentes des chercheurs et le travail des bibliothécaires et permettre la construction de projets concrets élaborés par des binômes chercheurs/bibliothécaires. L'axe principal de la collaboration entre CollEx-Persée et l'Abes s'inscrit dans cette logique. Cela se traduit par diverses actions, comme le projet de valorisation en ligne du patrimoine cartographique des bibliothèques CartoMundi, la fourniture de documents à distance à travers une plateforme collective mais surtout l'acquisition de corpus numériques sous licence nationale, auxquels l'Abes facilite l'accès en fournissant des métadonnées via la base de connaissance BACON ou en mettant en œuvre des opérations de signalement dédiés (cf p. 23)

MARIANNE GILOUX : j'ajouterai également notre participation à de nombreux appels à projets CollEx-Persée pour lesquels nous sommes régulièrement sollicités pour des opérations autour des données (fourniture de notices bibliographiques et d'autorité, enrichissement, alignement d'identifiants...) .

Comment s'articule cette collaboration avec les autres grands projets structurants de l'Abes ?

D. A. : c'est sans doute la grande question qui se pose dans tout le secteur de l'information scientifique et technique : comment les différents acteurs articulent-ils leurs actions ? Créé pour remplacer le réseau des Cadist et mieux intégrer les collections numériques au service des chercheurs, CollEx-Persée donne au paysage de l'IST française les moyens d'évoluer, grâce à une gouvernance adéquate et à des moyens d'action financiers et techniques. Cette impulsion induit, comme cela était souhaité dès l'origine, un repositionnement continu de chaque opérateur, processus complexe qui peut entraîner parfois des tensions mais que l'on résout par le dialogue.

M. G. : l'un des points structurants de notre projet d'établissement est le maintien du Sudoc au cœur de l'activité des bibliothèques, l'enrichissement de la qualité de ses données et l'amélioration des liens entre les données. Donc, répondre aux appels à projet de CollEx-Persée en fournissant des notices d'autorité et identifiants de chercheurs de la plateforme IdRef, participer avec le Sudoc-PS à la mise en œuvre des plans de conservation partagée des périodiques ou apporter notre expérience dans l'acquisition de ressources électroniques est tout à fait cohérent avec notre projet d'établissement.

Quelles sont les actions de l'Abes pour la valorisation des collections de CollEx-Persée ?

M. G. : l'Abes est copilote ou participe à plusieurs groupes de travail dédiés au signalement enrichi et à l'acquisition de ressources documentaires utiles aux chercheurs comme le projet de Cartographie des collections de CollEx-Persée (cf p.12). Notre équipe est aussi très impliquée dans le groupe de travail consacré aux acquisitions de ressources électroniques. Il s'agit de réaliser des achats ciblés de ressources souvent complexes dans leur structure et destinées à un public de spécialistes.

D. A. : la valorisation des collections labellisées se traduit par les nombreuses actions de l'Abes pour optimiser le signalement des ressources. Au-delà de la fourniture et de l'enrichissement de données bibliographiques ou d'autorité déjà évoqués, la synchronisation entre le Sudoc, les logiciels de bibliothèque, les bases de connaissance et les outils de découverte, ou encore l'exposition de nos données dans Google et les autres moteurs de recherche sont des opérations essentielles. Reste la question de l'accès à ces ressources. Personnellement, j'aimerais qu'il y ait une réflexion à l'échelle nationale sur les droits et conditions d'accès aux bibliothèques universitaires et je rêve parfois d'une carte de lecteur unique qui permettrait aux étudiants, aux enseignants, aux chercheurs d'entrer dans n'importe quelle BU et d'y emprunter n'importe quel document de ce que nous pourrions considérer comme la « collection nationale ».

Quelles sont les passerelles entre CollEx-Persée et la science ouverte ?

M. G. : différentes actions menées par l'Abes dans le cadre de CollEx-Persée contribuent à la science ouverte, comme par exemple le projet de diffusion nationale des thèses numérisées. Nous observons aussi que des laboratoires qui ont découvert les activités de l'Abes dans le cadre des appels à projets CollEx continuent ensuite de nous solliciter, nous plaçant ainsi au cœur de la science ouverte.

D. A. : les activités de CollEx-Persée qui consistent à acquérir, numériser, valoriser des collections nécessitent d'élaborer des métadonnées très fines pour exploiter ces ressources dans la logique de la science ouverte. Ainsi, le recours à l'identifiant pivot IdRef sert concrètement à la valorisation des chercheurs et de leurs productions en garantissant l'interopérabilité avec les systèmes d'identification nationaux (ScanR) et internationaux (ISNI, ORCID, VIAF, ORCID, Wikidata). Créées collectivement grâce au savoir-faire des bibliothèques de l'ESR, ces données d'une valeur inestimable facilitent l'accès aux ressources documentaires, accompagnant ainsi le grand virage de la science ouverte. ■

Fort de son expérience dans la numérisation et la diffusion des publications scientifiques, Persée a pris le rôle d'opérateur principal de CollEx-Persée dans la coordination de sa politique nationale de numérisation, apportant au réseau sa dimension d'infrastructure.

Persée : *numériser pour la recherche*

La numérisation, conçue non seulement comme l'opération de production de collections numériques mais également comme l'ensemble des services attachés à ces collections, constitue l'un des axes majeurs autour duquel s'est déployée l'action de CollEx-Persée. En effet, la numérisation des collections répond à deux grands enjeux qui sont également au cœur des missions du GIS. Le premier est celui d'une accessibilité large des matériaux et de la documentation pour la recherche : numériser pour diffuser, pour mieux signaler, pour reconstituer des séries complètes, pour préserver. Le second enjeu est celui de l'exploitation de ces ressources et de leur valorisation scientifique : numériser pour rendre exploitable par des outils numériques, pour enrichir les contenus, pour consolider les données. La numérisation enrichie promue par CollEx-Persée vise ainsi d'une part à numériser au plus près des problématiques des bibliothèques, qui ont vocation à assurer la disponibilité, la visibilité, la complétude et la conservation des collections, et d'autre part à numériser avec les chercheurs, pour répondre à leur besoin de corpus structurés et manipulables, enrichis par leur travail d'identification, de transcription, de balisage, de mise en relation avec les gisements de données pertinents dans un champ scientifique donné.

FAIRE CONVERGER LES PUBLICATIONS NUMÉRISÉES ET LES PUBLICATIONS NATIVEMENT NUMÉRIQUES

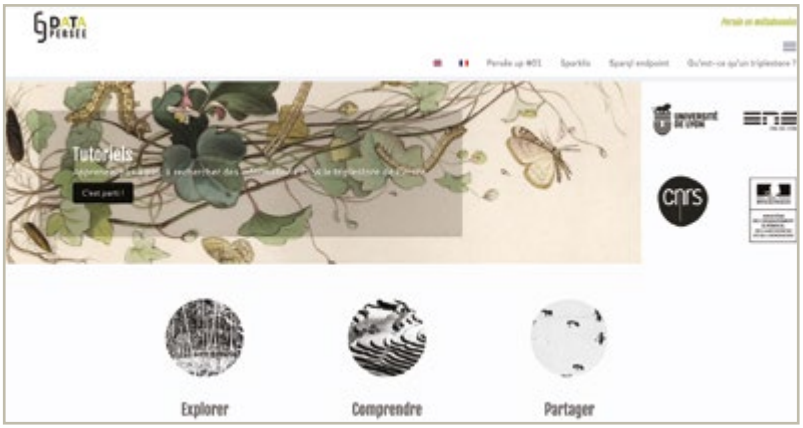
Depuis bientôt vingt ans, Persée numérise et diffuse les publications scientifiques –revues et collections de livres- françaises et francophones. Son portail[1] donne aujourd'hui accès à plus de 320 publications et collections, leur assurant une audience internationale, qui se traduit par 90 millions de visites annuelles en moyenne. Deux principes ont été, dès l'origine du programme, au fondement de l'activité de Persée : la production de contenus structurés et leur diffusion ouverte.

La structuration des contenus vise à diffuser les publications numérisées dans des formats qui convergent avec ceux de l'édition nativement numérique, afin d'y donner accès avec une granularité fine. L'enjeu est d'abord de faire de la numérisation des séries imprimées non seulement un vecteur de diffusion mais aussi le moyen de leur description fine et de leur enrichissement. Ce même objectif a déterminé les développements ultérieurs apportés à la production et à l'offre de service de Persée : le balisage d'entités nommées telles que les personnes, les taxons ou les lieux, leur alignement sur des référentiels extérieurs et tout d'abord sur IdRef ou la modélisation des métadonnées au format RDF et leur exposition dans le triplestore data.persée[2]. Tous ont été des étapes dans la consolidation de la qualité des données et dans la richesse des liens établis avec les autres sources d'information. Mais l'enjeu est aussi de répondre au besoin de préserver la continuité des publications qui font la science, en permettant que les publications antérieures soient aussi facilement accessibles, aussi bien référencées, aussi étroitement liées à l'environnement scientifique qui est le leur et qui est aujourd'hui massivement numérique, que les publications les plus récentes.

DES CORPUS ET DES OUTILS INSCRITS DANS UNE LOGIQUE *OPEN SOURCE*

Le principe d'ouverture tient à la fois à l'ouverture des données et à celle des outils. La mission initiale du programme d'assurer une diffusion libre du patrimoine scientifique français et francophone a évolué en même temps que s'est développée une réflexion internationale et qu'a été affirmée, au niveau national, une politique pour la science ouverte. Aujourd'hui, au-delà de l'adoption des standards qui visent à produire des données FAIR (faciles à trouver, accessibles, interopérables et réutilisables), Persée met à disposition ses corpus pour une exploitation scientifique et inscrit le développement de ses outils dans une logique *open source*. L'ouverture entraîne une exigence d'interopérabilité. L'objectif d'une diffusion ouverte, comme celui d'une réutilisation des données en dehors du contexte initial de leur

[1] https://www.persee.fr
[2] http://data.persee.fr



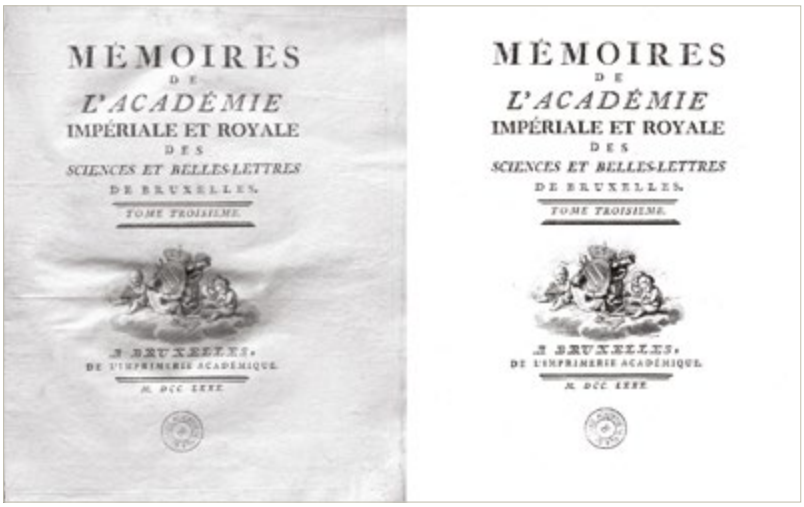
production ne peut être atteint qu'à la condition que les plateformes sur lesquelles elles sont produites, traitées et hébergées offrent les moyens non seulement de les consulter mais aussi de les manipuler. C'est dans ce but que le protocole IIIF a été implémenté sur la nouvelle plateforme de Persée ou que les corpus de Persée sont mis à disposition d'équipes de recherche pour des opérations de traitement automatique des langues et de fouille.

PRODUIRE DES CORPUS EN LIEN AVEC DES PROJETS DE RECHERCHE

Ces choix initiaux et les développements qu'ils ont conduits à mettre en œuvre ont amené Persée, depuis 2018, à élargir son activité à la production de corpus en lien avec des projets de recherche. Le périmètre documentaire des documents traités s'est ainsi élargi à d'autres sources. Ces corpus, diffusés sur des sites Web dédiés appelés Perséides, incluent aujourd'hui, suivant les projets, des publications administratives, des albums iconographiques, des éditions de sources et bientôt des carnets de fouille manuscrits. Le modèle de structuration et la nature des enrichissements, spécifiques à chaque corpus, sont définis conjointement par les équipes de Persée et les équipes de recherche qui portent le projet.
Persée a donc acquis, à travers ces années de développement et de production, une expérience singulière dans la numérisation. Singulière parce qu'elle s'appuie sur une chaîne intégrée actuellement sans équivalent, qui prend en charge l'intégralité des opérations depuis l'acquisition des images jusqu'à la diffusion de données structurées. Singulière aussi parce qu'elle conjugue un niveau élevé d'enrichissements et une volumétrie de production qui, même si elle ne peut répondre à l'ampleur avérée des besoins – ceux des équipes de recherche qui souhaitent disposer des sources documentaires nécessaires à leurs travaux, ceux des bibliothèques qui veulent rendre plus accessibles leurs collections et ceux des éditeurs qui cherchent à valoriser leurs publications – fait néanmoins de Persée l'un des acteurs majeurs de la numérisation.

EN APPUI AU RÉSEAU, DANS UNE LOGIQUE D'INFRASTRUCTURE

Fort de cette expérience, Persée a pris le rôle d'opérateur principal de CollEx-Persée, avec la charge de coordonner la politique nationale de numérisation portée par le GIS, apportant au réseau sa dimension d'infrastructure. De nombreux travaux ont été menés dans le cadre de cette politique : travaux d'analyse et d'enquête conduits par le groupe Numérisation de CollEx-Persée[3] pour un état des lieux de la numérisation dans l'enseignement supérieur et la recherche, appui à l'exploration des fonds à travers l'appel à résidences, enfin soutien décisif à des projets de numérisation enrichie apporté par les trois éditions



de l'appel à projets « Numérisation et services ». La politique de numérisation a également permis la mise en œuvre d'un programme de numérisation concertée entre Persée et les plans thématiques de conservation partagée des périodiques (PCPP)[4]. Avec l'ambition de conjuguer les objectifs des PCPP, qui portent sur l'identification, le signalement et la conservation des périodiques, et des objectifs de large diffusion et de référencement à l'article grâce à la numérisation, ce programme s'appuie à la fois sur l'expertise scientifique et documentaire développée au sein des PCPP pour la constitution des corpus de titres, sur la constitution des PCPP en réseaux, qui facilite la mise à disposition d'une collection complète, et sur la chaîne et l'équipe de Persée pour la numérisation et l'enrichissement des contenus. Plus de 230 000 pages ont été produites depuis le début du programme en 2020, issues principalement de titres des PCPP Antiquité et Psychologie.
Les travaux conduits dans le cadre de ce GIS fondent aujourd'hui les ambitions d'un nouveau programme de numérisation pour le GIS CollEx-Persée 2. La trajectoire entamée invite à le concevoir avec trois objectifs majeurs : soutenir les projets exploratoires de numérisation enrichie qui ont permis d'accompagner des projets de recherche, consolider l'apport du GIS en termes d'infrastructure par la réalisation d'une plateforme de numérisation ouverte à l'ensemble du réseau, porter des projets de numérisation de grande ampleur pour répondre aux besoins de la communauté scientifique de disponibilité des matériaux pour la recherche et des collections.

Gabrielle Richard
Directrice de Persée
gabrielle.richard@persee.fr

[3] Pour plus d'informations sur les travaux du GT Numérisation, voir https://www.collexpersee.eu/groupes/numerisation

[4] Voir https://www.collexpersee.eu/diffusion-ouverte-et-preservation-des-collections-de-periodiques-pour-unprogramme-national-de-numerisationconcertee-persee-pcp

Le 25 novembre 2021, une journée d'étude a fait le bilan du projet Cartographie CollEx-Persée, destiné à valoriser les collections labellisées « Collections d'Excellence », et a esquissé les perspectives d'avenir.

Cartographie CollEx-Persée : *valoriser les « Collections d'Excellence »*

Le 25 novembre 2021, La contemporaine n'avait pas encore fêté l'inauguration officielle de ses nouveaux locaux qu'elle accueillait déjà une journée d'étude consacrée à la Cartographie CollEx-Persée, projet destiné à valoriser les collections labellisées « Collections d'Excellence ». Quarante-cinq personnes, représentant les seize établissements impliqués à ce jour dans le projet, s'étaient rassemblées pour discuter des enjeux et des réalisations de l'année 2021. Elles devaient aussi, collectivement, imaginer l'avenir de ce nouvel outil de signalement en réseau.

DÉCRIRE UNE RÉALITÉ COMPOSITE AU SEIN DU RÉPERTOIRE DE FONDS DU CATALOGUE COLLECTIF DE FRANCE

Le projet d'une cartographie dynamique, pour favoriser l'usage des collections des bibliothèques par les chercheurs, figure dans la feuille de route initiale du GIS CollEx-Persée. Il s'agissait de souligner l'intérêt de gisements documentaires largement méconnus au sein des bibliothèques, dans un contexte où la disponibilité massive de ressources documentaires en ligne a modifié les pratiques de recherche. S'appuyant sur les bases de signalement existantes, sans chercher à les concurrencer, cet outil entend répondre à des exigences liées au caractère très hétérogène des ressources à décrire. Généraliste et pluridisciplinaire, il couvre l'ensemble du spectre académique, mais aussi tous les types de documents susceptibles d'être conservés : imprimés, manuscrits, archives, fonds iconographiques, collections numériques.

Après une première phase d'étude en 2019, il a été décidé de s'appuyer sur le répertoire de fonds du Catalogue collectif de France (CCFr)[1]. Il offre, en effet, l'avantage de concerner l'ensemble des bibliothèques ouvertes au public, relevant ou non de l'Enseignement supérieur et de la recherche. En outre, l'évolution de cet outil vers une solution en EAD, à l'orée 2022, ouvre des perspectives d'interopérabilité avec d'autres plateformes (Calames, IdRef), ainsi que la possibilité pour les chercheurs d'exporter et de réutiliser les données, celles-ci étant sous licence ouverte Etalab.

TESTER UN MODÈLE DE DONNÉES QUI VALORISE LES THÉMATIQUES DES CORPUS DOCUMENTAIRES

En 2020, un modèle de données a été élaboré, qui articule trois niveaux : institution détentrice, fonds et sous-fonds. L'enjeu est de rendre visibles, dès l'énoncé du titre et du sous-titre, les thématiques, la provenance et les supports des ensembles signalés. Ces informations doivent apparaître d'emblée dans le moteur de recherche du répertoire pour faciliter la navigation de l'usager chercheur. Afin de tester le modèle de données proposé et de fournir un ensemble conséquent d'exemples, un groupe pilote a été constitué début 2021, représentatif de divers types de bibliothèques et de disciplines[2]. Très vite, les établissements se sont pris au jeu et ce sont aujourd'hui près de 550 notices « CollEx-Persée » qui sont mises à disposition. Couvrant l'ensemble des domaines scientifiques du référentiel Hcéres (Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur), ces descriptions démontrent la pertinence de l'approche retenue : elles serviront de modèles pour l'ouverture du service à l'ensemble des collections labellisées.

Lors de la journée d'étude du 25 novembre, les pilotes du projet, La contemporaine, le Muséum national d'histoire naturelle, la Bibliothèque nationale de France et l'Agence bibliographique de l'enseignement supérieur se sont succédés pour proposer le bilan de cette première phase et préfigurer les prochaines étapes : objectifs, gouvernance et développements informatiques. L'après-midi était plus interactive : réunis en différents ateliers thématiques (Sciences de l'Antiquité, Histoire politique contemporaine,

[1] https://ccfr.bnf.fr/portailccfr/jsp/public/index.jsp?action=public_formsearch_fonds

[2] Le groupe pilote était constitué des bibliothèques de l'INHA (Institut national d'histoire de l'art), de Sciences-Po, du MNHN, de Sorbonne Université, de la Sorbonne Nouvelle, de l'Université Grenoble Alpes et de la bibliothèque Henri-Piéron de l'Université de Paris, des SCD des universités de Lille et de Lyon 1, de la médiathèque du Musée du Quai Branly, des bibliothèques interuniversitaires de Santé et de la Sorbonne, de la Bibliothèque nationale et universitaire (BNU), de La contemporaine, de la Bibliothèque universitaire des langues et civilisations (Bulac) et du GED Condorcet.

© Haller von Hallerstein Collection Bnu

Dessin à l'aquarelle du temple de Thésée, à Athènes, sans date : vers 1810-1817.

Anthropologie et Ethnologie, Sciences et techniques, Atelier général sur la Cartographie), les participants ont pu réfléchir aux perspectives d'avenir de la Cartographie, qui a vocation à s'enrichir des travaux menés par les communautés de chercheurs.

S'APPUYER SUR LES PROJETS DE RECHERCHE POUR TRAVAILLER EN RÉSEAU

À l'issue de la journée, les restitutions ont permis de constater de nombreuses convergences, quelle que soit la thématique considérée. L'atelier « Anthropologie et Ethnologie » a souligné que l'exercice de cartographie était une étape préalable à de nombreux de projets de recherche qui s'emploient souvent, dans leur phase initiale, à recenser les matériaux disponibles. Adoptant une métaphore naturaliste, on pourrait comparer la Cartographie à un épiphyte, cette plante qui s'appuie sur les autres pour monter vers le ciel. Chaque travail de recensement et d'indexation mené dans le cadre d'un projet de recherche viendrait enrichir le répertoire de fonds du CCFr tandis que cet outil, plus pérenne, jouerait un rôle de pivot pour faire dialoguer les fonds ainsi décrits avec d'autres ressources. Enfin, le répertoire offrirait aux chercheurs de nombreuses possibilités d'exports et de réutilisations.

Ces enjeux expliquent les interactions étroites entre la Cartographie et le projet « Digital Maghreb [3] », porté par le GIS « Moyen-Orient et mondes musulmans ». Cet organisme a diligenté une enquête pour identifier les fonds intéressant les études maghrébines dans les collections françaises. Celle-ci recense des fonds déjà signalés dans le répertoire du CCFr par des établissements labellisés CollEx, tout en apportant un éclairage thématique. La Cartographie joue alors son rôle de mutualisation des projets dans la mesure où une indexation spécifique à un axe de recherche peut enrichir des notices existantes. L'atelier « Sciences de l'Antiquité » a justement souligné le besoin de répondre, dans les pratiques de description, aux approches des chercheurs. Ainsi, les fonds relevant de cette thématique devraient, en plus de Rameau, adopter le thésaurus PACTOLS, conçu par et pour des chercheurs, et utilisé par des collections conservées dans des établissements relevant du ministère de la Culture.

QUEL CADRE DE TRAVAIL POUR LA CARTOGRAPHIE EN 2022 ?

Outre la mise en place d'ateliers thématiques, l'enjeu de l'année 2022 est d'étendre la Cartographie au plus grand nombre d'établissements possédant des collections labellisées. À cette fin, un webinaire de présentation sera organisé au premier trimestre. En parallèle, le « groupe de suivi » Cartographie sera maintenu pour accompagner les collègues débutant leurs travaux de signalement.

© Collection Bnu – Numistral

Andreas Cellarius, Scenographia compagis mundanæ Brahea, 1708.

En matière de description, deux types d'ensembles ont été identifiés comme prioritaires :

- Les fonds dits « dormants », car non signalés dans les catalogues et pour lesquels le répertoire du CCFr offre une solution adaptée pour attirer l'attention des chercheurs.
- Les fonds hybrides croisant matériaux physiques et numériques, imprimés, archives et objets, déjà signalés par ailleurs mais répartis sur plusieurs catalogues. Les notices de fonds du CCFr permettraient de faire apparaître une cohérence dont ne rendent pas compte des catalogues spécifiques. Cette même logique peut également permettre de reconstituer l'unité originelle d'ensembles aujourd'hui conservés par différents établissements.

Un autre enjeu sera de veiller aux bonnes pratiques de signalement à travers un travail normatif. En 2021, le chef de projet Cartographie a ainsi utilisé le moteur de recherche ScanR pour identifier les points d'accès employés de préférence par les communautés de recherche susceptibles de mobiliser les collections décrites. Chaque réseau thématique pourrait capitaliser sur ces travaux afin d'adapter les indexations retenues aux pratiques des utilisateurs. Enfin, une attention particulière sera portée à la mise à jour régulière des notices produites. Ce suivi indispensable, qui doit être porté par la demande et l'usage des chercheurs, garantira le dynamisme de la Cartographie et permettra d'assurer sa pérennité dans un projet CollEx-Persée 2.

Clément Oury
Copilote du projet Cartographie CollEx-Persée, directeur adjoint de la bibliothèque du Muséum national d'histoire naturelle
clement.oury@mnhn.fr

Philippe Goguely
Chef du projet Cartographie CollEx-Persée, bibliothèque du Muséum national d'histoire naturelle
phlippe.goguely@mnhn.fr

[3] https://digimagh.hypotheses.org

Datacc : une approche renouvelée de la notion de données de et pour la Recherche

RETOUR D'EXPÉRIENCE

Le projet Datacc accompagne les chercheurs en chimie et physique à la gestion de leurs données et les sensibilise aux solutions permettant d'améliorer la traçabilité des travaux scientifiques, leur reproductibilité et l'efficacité de la recherche.

L'implication dans le dispositif CollEx-Persée d'un gros service commun de la documentation (SCD) scientifique comme celui de Lyon 1 permet d'apporter un éclairage complémentaire à celui des SHS sur les besoins et pratiques des chercheurs et de faire émerger des projets de services innovants, sur la gestion des données en particulier, qui renouvellent les notions de corpus et matériaux pour la recherche.

L'expérience acquise par Lyon 1 en tant que Cadist (Centre d'acquisition et de diffusion de l'information scientifique et technique) – en chimie et pharmacie – et les efforts menés tout au long des 20 dernières années pour s'adapter aux évolutions du paysage informationnel de l'enseignement supérieur et de la recherche lui permettent de répondre à la double ambition du GIS CollEx-Persée pour les établissements du réseau.

En premier lieu, il répond à l'ambition de bibliothèques qui développent des collections d'excellence et en facilitent l'accès et l'usage par les chercheurs. En rassemblant en quelques pôles de conservation identifiables des collections de revues auparavant partielles et dispersées, le Plan national de conservation partagée des périodiques en Chimie, déployé au sein d'un réseau de 23 bibliothèques piloté par le SCD Lyon 1, offre une carte documentaire beaucoup plus cohérente et lisible pour les chercheurs.

Articulé aux collections d'excellence labellisées, le service de fourniture de documents à distance du SCD Lyon 1 a été récemment modernisé afin d'améliorer leur référencement et d'en optimiser l'accès. Le déploiement d'une solution de fourniture dématérialisée de documents, acquise en 2018 sur des crédits CollEx-Persée, permet une transmission électronique dans la journée de l'article demandé par le chercheur. L'obtention du document est d'autant plus aisée et rapide qu'en tant que membre de *WorldShare Interlibrary Loan,* le service peut se fournir auprès de plus de 10 000 bibliothèques, en même temps qu'il dessert les communautés scientifiques des 63 pays du réseau. L'acquisition de ressources électroniques de niche en licence nationale représente un autre volet de cette stratégie de développement de collections hybrides aisément accessibles aux chercheurs, auquel le SCD Lyon 1 a récemment contribué en tant que responsable scientifique du projet d'achat des archives de revues en chimie chez *Bentham Science Publisher.*

TROUVER LES SOLUTIONS ÉLECTRONIQUES POUR UNE MEILLEURE GESTION DES DONNÉES DE LA RECHERCHE

En second lieu, la vision de CollEx-Persée est celle de bibliothèques développant des services documentaires répondant aux nouveaux besoins des chercheurs en tant qu'utilisateurs d'information scientifique et producteurs de données et de résultats de recherche. L'élaboration du projet Datacc en 2019 en est le reflet : au-delà de l'accompagnement à la rédaction de plans de gestion de données, problématique de plus en plus prise en compte par les services à la recherche des bibliothèques universitaires, le projet s'est ancré autour de la sensibilisation des chercheurs aux outils scientifiques permettant le traitement, l'analyse et la mise en forme des données, avant leur potentielle ouverture.

Les expériences sont encore, le plus souvent, décrites dans un cahier de laboratoire papier, dont la nature même du support ne permet pas une réutilisation optimale du contenu. L'un des axes centraux du projet a donc consisté à repérer les solutions électroniques assurant une meilleure préservation de ce savoir-faire scientifique, dont le contenu détaillé ne figure pas dans les publications académiques. Un accompagnement est proposé aux chercheurs pour organiser leur cahier de laboratoire à l'échelle de leur équipe au sein d'une instance numérique partagée.

Un signalement des entrepôts disciplinaires en chimie et en physique a également été publié sur le site Datacc.org, accompagné de notices permettant de mieux cerner la politique de chaque plateforme. Enfin, une réflexion est menée sur les biais de publication conduisant à l'absence de partage des données issues d'expériences scientifiques infructueuses, communément désignées sous le vocable de « résultats négatifs » Une enquête[1] a été menée en ce sens afin d'identifier, notamment, l'intérêt scientifique et l'appétence des communautés pour la création d'une plateforme consacrée à la diffusion de ce type de résultats.

Amanda Cruguel
Cheffe du département Services à la recherche, SCD de l'université Claude Bernard Lyon 1
amanda.cruguel-vitry@univ-lyon1.fr

Marie-Emilia Herbet
Cheffe de projet Datacc, SCD de l'université Claude Bernard Lyon 1
marie-emilia.herbet@univ-lyon1.fr

DATACC

Université Claude Bernard Lyon 1

[1] https://hal.archives-ouvertes.fr/hal-03371040

Le projet NumBois, lauréat de l'appel à projets CollEx-Persée en 2020, permettra de numériser 3 000 échantillons de bois de la Xylothèque du Cirad et de les enrichir en métadonnées.

NumBois, rencontre entre la recherche en sciences du bois et l'information scientifique et technique

RETOUR D'EXPÉRIENCE

Lauréat de l'appel à projets CollEx-Persée en 2020, le projet NumBois[1] est porté par deux collectifs du Cirad, la Dist (Délégation à l'information scientifique et technique) et l'équipe de recherche *BioWooEB* en partenariat avec des universités françaises et africaines et des professionnels du bois. NumBois s'articule autour de trois initiatives et corpus existants : NumBA, la bibliothèque numérique du Cirad en agronomie tropicale[2], la Xylothèque du Cirad[3], une des plus importantes collections d'échantillons de bois tropicaux et tempérés dans le monde et la base de données du Cirad sur les propriétés technologiques des bois[4] associée à l'application Tropix version 7[5]. Lancée en 2019, NumBA est la bibliothèque numérique du Cirad conçue en Gallica Marque blanche avec la Bibliothèque nationale de France (BnF). La collection patrimoniale « De l'agronomie coloniale à l'agronomie tropicale : histoire de la recherche française depuis le XIXe siècle », labellisée CollEx en décembre 2017, est le fondement de cette bibliothèque. NumBA propose près de 9 000 documents en ligne, en accès libre et gratuit. Elle s'enrichit régulièrement avec de nouvelles collections textuelles, iconographiques et des objets originaux d'origine végétale.

Par ailleurs, le Cirad possède une xylothèque dont l'origine remonte à 1937. Aujourd'hui consultable sur place uniquement, elle rassemble plus de 34 000 spécimens de bois tropicaux et tempérés correspondant à 8 400 espèces, 2 160 genres et 235 familles botaniques. Elle conserve des essences provenant de 123 pays d'Afrique, d'Amérique, d'Asie, d'Océanie et d'Europe.

Source : Xylothèque du Cirad

Spécimens de bois tropicaux de la Xylothèque.

Complémentaire de la Xylothèque, la base de données et de connaissances sur les propriétés technologiques des bois décrit plus de 1 300 espèces tropicales et méditerranéennes. Les données relatives à 245 de ces essences sont accessibles via le logiciel propriétaire Tropix. L'ensemble constitué par la Xylothèque, la base de données et le logiciel Tropix représente un patrimoine scientifique et culturel remarquable sur le bois et constitue un outil de recherche, d'expertise et de formation reconnu des spécialistes.

LA CONSTRUCTION D'UN PROJET COLLABORATIF

L'objectif du projet NumBois est de valoriser les résultats et les connaissances capitalisées, d'en faciliter l'accès et d'en élargir les usages pour la recherche, pour les professionnels de la filière bois et pour un public averti. Le soutien de CollEx-Persée a été le déclencheur et un levier pour le projet. Il permettra de numériser 3 000 échantillons de bois de la Xylothèque du Cirad, de les enrichir en métadonnées et de les intégrer dans la bibliothèque numérique du Cirad, NumBA. En parallèle, le logiciel Tropix d'identification des caractéristiques technologiques de bois tropicaux sera déployé sur Internet en relation avec NumBA. L'ensemble des données et des documents sera en accès libre et gratuit, conformément au principe du libre accès aux connaissances inscrit dans la stratégie institutionnelle du Cirad depuis 2006.

CollEx-Persée a donné une impulsion au projet NumBois en encourageant la collaboration d'une équipe de recherche du Cirad, de ses partenaires en France et à l'international, et de la délégation à l'information scientifique et technique du Cirad. Le projet bénéficie ainsi de l'expertise conjointe de tous les acteurs : les chercheurs produisent les données et les connaissances scientifiques ; les professionnels de l'IST structurent ces données et enrichissent leurs métadonnées selon les référentiels de la BnF pour les rendre accessibles au plus grand nombre dans NumBA.

En soutenant le projet NumBois, le GIS CollEx-Persée accompagne ainsi le Cirad dans ses efforts de valorisation et de diffusion du patrimoine scientifique institutionnel.

Sylvie Vago
Responsable des bibliothèques du Cirad
sylvie.vago@cirad.fr

Marie-Françoise Fily
Documentaliste à la Délégation à l'information scientifique et technique du Cirad
marie-francoise.fily@cirad.fr

[1] https://www.collexpersee.eu/projet/numbois

[2] https://numba.cirad.fr/numba/?mode=desktop

[3] https://ur-biowooeb.cirad.fr/expertises-produits/expertises-bois-materiau/xylotheque

[4] https://ur-biowooeb.cirad.fr/expertises-produits/expertises-bois-materiau/bases-de-donnees

[5] https://tropix.cirad.fr

Rassemblant 20 000 lettres numérisées, FamiliLettres, porté par le Cnam et plusieurs partenaires avec l'aide de CollEx-Persée, est à la fois un projet de valorisation patrimoniale, d'édition et un terrain d'expérimentation.

FamiliLettres : un accès unique à un corpus de correspondances exceptionnel

RETOUR D'EXPÉRIENCE

FamiliLettres est le projet d'édition sur la plateforme EMAN des correspondances privées, actives et passives, 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 sa collaboratrice et compagne Marie Moret (1840-1908). Il est porté par la Direction des bibliothèques et de la documentation du Cnam (Conservatoire national des arts et métiers) et le Familistère de Guise, qui conservent ces correspondances, avec deux laboratoires du Cnam, le Lise (Laboratoire interdisciplinaire pour la sociologie économique, UMR CNRS-Cnam n° 3320) et le Cédric (Centre d'études et de recherche en informatique et communications, EA4629).

L'idée de l'édition de ce corpus de près de 20 000 lettres est née fin 2016 lors d'une rencontre entre la bibliothèque centrale du Cnam et le Familistère, l'objectif étant de donner un accès unique en ligne à cet ensemble physiquement dispersé, à la suite des inquiétudes quant à sa conservation et présentant des difficultés de lecture. En effet, ce corpus est exceptionnel à bien des égards, par sa volumétrie, sa complétude et sa richesse. Les deux auteurs ont ainsi pris soin de conserver de manière systématique dans des registres une copie des lettres qu'ils envoyaient, copie réalisée par le moyen d'une presse à copier sur une période très longue entre 1843 et 1908. Il constitue un des rares exemples de double correspondance active et passive, bien que cette dernière soit beaucoup plus lacunaire, pour l'histoire sociale et politique du XIXe siècle. Au-delà du fouriérisme, il intéresse de nombreux champs de recherche : histoire de l'éducation, du féminisme, de la vie privée, de l'architecture du logement, du pacifisme, du mouvement spirite, histoire et sociologie des communautés intentionnelles, sociologie des *prefigurative politics*[1].

Mais la réalisation de ce projet avait alors achoppé sur le manque de moyens humains et financiers et le constat que les outils à notre disposition, et notamment la bibliothèque numérique du Cnam, le Cnum, développée à l'origine pour valoriser un corpus de livres imprimés, n'offrait pas les fonctionnalités nécessaires à une telle entreprise (schéma de métadonnées contraint, impossibilité de lier des notices entre elles). La recherche d'un tel outil nous a amenés en 2019 à nous rapprocher de la plateforme EMAN[2], outil de publication numérique basé sur Omeka pour la diffusion et l'exploitation de manuscrits et de fonds d'archives modernes qui permet de diffuser à la fois les documents numérisés, leurs descriptions, leurs transcriptions et leurs analyses scientifiques et qui hébergeait déjà plusieurs projets d'édition de correspondances.

LE FINANCEMENT DE COLLEX-PERSÉE : UN LEVIER INDISPENSABLE

Le financement de CollEx-Persée a permis le recrutement d'un ingénieur d'études durant 24 mois, sur qui reposent notamment l'enrichissement des métadonnées descriptives de la correspondance (auteur, destinataire, lieu de rédaction et d'expédition mais aussi œuvres citées, personnes citées, événements cités, résumé…), l'enrichissement des métadonnées relatives aux correspondants (plus de 3 100 identifiés à ce jour), la transcription des lettres et l'administration de notre site sur EMAN. Il nous a permis aussi d'obtenir les crédits nécessaires pour soutenir les évolutions de la plateforme EMAN à travers de nouveaux développements informatiques, dont profiteront les autres projets hébergés sur EMAN. Enfin, avec le Cédric et l'appui d'une stagiaire de master 2, nous avons travaillé sur un volet plus expérimental pour la réalisation d'une interface permettant la visualisation des données et facilitant la compréhension et la déambulation dans le corpus.

Bien que l'investissement de l'équipe dans FamiliLettres se fasse parfois au détriment d'autres activités, ce projet, qui est tout à la fois un projet de valorisation patrimoniale, d'édition et un terrain d'expérimentation, a permis, outre une montée en compétence dans le domaine des humanités numériques, de créer une synergie de travail avec plusieurs laboratoires de recherche et de mieux connaître nos collections pour lesquelles nous espérons une visibilité et une exploitation accrues.

Cécile Formaglio
Responsable du pôle Patrimoine et conservation à la direction des bibliothèques et de la documentation du Cnam
cecile.formaglio@lecnam.net

[1] Les politiques préfiguratives sont les modes d'organisation et de relations sociales qui s'efforcent de refléter la société future recherchée par le groupe.

[2] https://eman-archives.org

le cnam

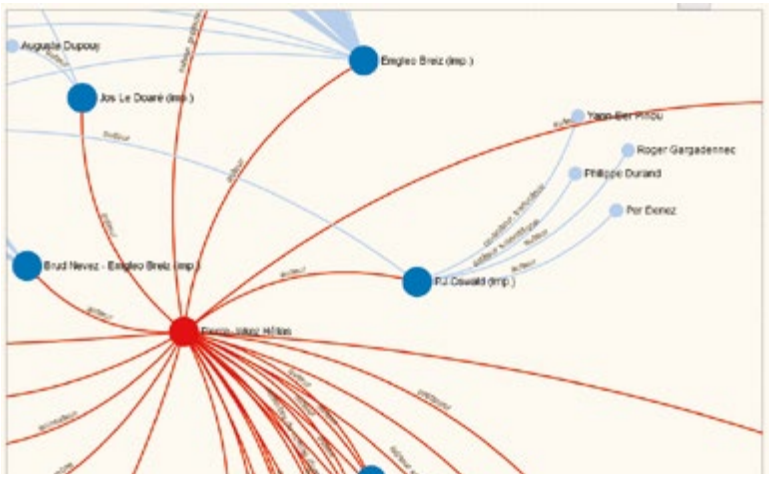
Ce projet de l'université de Bretagne occidentale baptisé REBELLE, portant sur la Fondation culturelle bretonne, est le résultat d'un travail commun entre chercheurs et personnels d'appui à la recherche.

La République locale des lettres

Analyse du jeu littéraire et militant en Bretagne

RETOUR D'EXPÉRIENCE

Le programme de recherche « La République locale des lettres » (REBELLE), porté par le Centre de recherche bretonne et celtique (CRBC) de l'université de Bretagne occidentale, à Brest, a été lauréat en 2019 d'un appel à projets CollEx-Persée Numérisation. Il a pour but d'analyser la manière dont a émergé l'un des acteurs majeurs du domaine littéraire et militant en Bretagne au début des années 1950 : la Fondation culturelle bretonne [*Emgleo Breiz*]. Ce programme de recherche vise à rendre accessibles aux chercheurs de multiples documents sur les activités de ses membres, mais aussi de faciliter l'accès et le traitement des données qui en sont issues afin de mener une analyse approfondie de l'activité littéraire et militante en Bretagne entre 1946 et 1977. Ceci n'est pas uniquement d'intérêt local ou régional : les individus concernés entretenaient de nombreuses relations avec des personnalités hors de Bretagne, au point de constituer un véritable réseau interrégional sur des questions culturelle, politique et de promotion de l'enseignement des langues régionales. Si une partie des archives est conservée à la bibliothèque Yves Le Gallo (UMS 3554 CRBC-Documentation CNRS/UBO), d'autres partenaires ont été impliqués parce qu'ils disposent de fonds permettant d'enrichir ces ressources initiales : les Archives départementales du Finistère et la médiathèque occitane du Cirdoc (Centre international de recherche et de documentation occitanes).



© UMS CRBC-Documentation

Réseau de sociabilités de l'auteur Pierre-Jakez Hélias (1914-1995).

UN TRAVAIL DE SYNERGIE ET QUI S'INSCRIT DANS LA DYNAMIQUE DE LA SCIENCE OUVERTE

Ce programme a été mené en synergie entre les institutions impliquées (institutions de conservation patrimoniale et laboratoires) de même qu'entre personnels d'appui à la recherche et chercheurs du CRBC et du laboratoire RESO de l'université Paul-Valéry Montpellier 3, selon les étapes suivantes :

- Normalisation archivistique et structuration des données.
- Alignement avec les notices d'autorité d'IdRef et la base de données PRELIB[1].
- Numérisation et mise en ligne d'une sélection de plus de 550 documents sur la bibliothèque numérique du CRBC[2] et près de 450 sur celle du Cirdoc[3], accompagnées de billets de contextualisation sur le carnet Hypothèses dédié[4].
- Mise à disposition des inventaires et des données sur différentes plateformes et portails dont CRBC Dataset[5] et FranceArchives[6].
- Des ateliers et journées d'études réunissant des chercheurs spécialistes de ces différentes questions qui ont permis d'analyser les revendications linguistiques portées au niveau régional puis interrégional par Armand Keravel (1910-1999) – principale cheville ouvrière d'*Emgleo Breiz* – et par Robert Lafont (1923-2009), l'une des figures majeures du mouvement occitan. Leur correspondance, notamment, a permis de documenter la simultanéité de certains phénomènes dans les différents mouvements militants régionaux, mais aussi de comprendre la mise en œuvre d'une véritable stratégie interrégionale en faveur de l'enseignement des langues régionales de France.

S'inscrivant dans la dynamique de la science ouverte développée au sein du CRBC depuis plusieurs années (bibliothèque numérique, revues en *open access*, développement du Web sémantique), le programme de recherche REBELLE a nécessité un travail commun entre chercheurs et personnels d'appui à la recherche. Une correspondance aussi volumineuse a nécessité, d'une part l'enrichissement des métadonnées et d'autre part la numérisation, la structuration et la mise en ligne de ces documents. L'ensemble de ce travail commun en a grandement facilité l'accès et l'usage. Dernière perspective dans le cadre de REBELLE : un projet de mise en réseau des centres de documentation et d'archives Langues de France a été lancé en novembre 2021 avec le Cirdoc.

Mannaig Thomas
Maître de conférences à l'université de Bretagne occidentale
mannaig.thomas@univ-brest.fr

Philippe Lagadec
Ingénieur d'études CRBC-UBO, pour REBELLE-CollEx
philippe.lagadec@univ-brest.fr

UBO
Université de Bretagne Occidentale

[1] https://mshb.huma-num.fr/prelib
[2] https://bibnumcrbc.huma-num.fr
[3] https://occitanica.eu
[4] https://rebelle.hypotheses.org
[5] https://crbc-dataset.huma-num.fr
[6] https://francearchives.fr/fr

Le projet collaboratif Mistara réactualise les problématiques d'alignement et de qualité des données en écriture non latine tout en les confrontant aux nouveaux outils de l'Abes et de la BnF.

Mistara : au croisement des enjeux de la transition bibliographique et de la recherche sur l'onomastique - ou étude des noms - arabo-musulmane

RETOUR D'EXPÉRIENCE

La mistara, tablette de réglure utilisée pour tracer les lignes d'écritures dans les manuscrits du monde islamique, est la métaphore parfaite pour un projet centré autour de l'alignement des noms de personnes en écriture arabe. Le projet Mistara[1], financé par le GIS CollEx-Persée de 2018 à 2021, a été porté par la Bulac (Bibliothèque universitaire des langues et civilisations) en partenariat avec le service commun de la documentation d'Aix-Marseille Université, la section arabe de l'Institut de recherche sur l'histoire des textes (CNRS), l'Institut dominicain d'études orientales du Caire, la bibliothèque de l'Institut du monde arabe et la Maison méditerranéenne des sciences de l'Homme, avec le soutien des agences bibliographiques nationales. La vertu première de ce type de financement a été d'encourager une démarche collaborative. Elle a permis de réactualiser et de traiter des problématiques déjà connues d'alignement et de qualité des données en écriture non latine, tout en les confrontant aux nouveaux outils élaborés par l'Abes et la BnF dans le cadre de la transition bibliographique. Le projet visait à utiliser des référentiels tiers issus de la recherche : la base AlKindi de l'Institut dominicain d'études orientales [Ideo] et l'*Onomasticon Arabicum* de l'Institut de recherche et d'histoire des textes [IRHT] pour enrichir IdRef.

Le projet croise les enjeux de la transition bibliographique et de la recherche sur l'onomastique arabo-musulmane, en cherchant d'une part à évaluer les adaptations nécessaires à la modélisation et à la manipulation des formats d'entités personne, à la gestion d'entités multilingues et multiécritures et aux spécificités de la décomposition du nom dans le monde arabo-musulman, et d'autre part à améliorer la gestion des référentiels produits par la recherche sur les questions onomastiques du monde arabo-musulman pour assurer leur interopérabilité et permettre l'enrichissement des outils catalographiques.

UN TRAVAIL COOPÉRATIF ENTRE PARTENAIRES ISSUS DE DIFFÉRENTS HORIZONS

Les objectifs du projet ont été atteints en bonne partie. Les données des différents partenaires ont permis d'enrichir les données du Sudoc et de tester leur traitement avec différents outils d'alignement ou de correction des liens entre notices : Bibliostratus, Openrefine, Qualinka/Paprika. L'hétérogénéité des données de départ a tout d'abord nécessité un traitement humain avant de pouvoir utiliser les outils existants. Un rapport détaillé sur le bilan d'usage de ces outils pour des données en écriture arabe et des données décrivant des entités physiques a été élaboré avec l'Abes. Les consignes de traitement des entités en caractères arabes ont pu être précisées, les bases d'exemples enrichies pour préciser et homogénéiser la prise en compte des pratiques d'écriture et la richesse onomastique du monde islamique : ces informations seront mises à disposition du réseau de l'Abes sous la forme d'un guide complémentaire pour le catalogage des autorités arabo-musulmanes.

La dimension coopérative du travail autour de l'alignement des référentiels s'est avérée particulièrement riche. Une dynamique nouvelle a pu se créer entre des partenaires dans et hors du champ de l'enseignement supérieur, et des opportunités intéressantes ont surgi. Une collaboration est entamée entre l'Ideo et l'Abes qui devrait permettre aux établissements qui utilisent l'environnement *ILS-DIAMOND* [nativement LRM] de produire leurs données dans IdRef et, à terme, d'enrichir rétrospectivement le référentiel avec les notices produites jusqu'à ce jour. Ce sera un enrichissement essentiel du point de vue de l'onomastique arabe qui profitera également aux études de la patristique latine et orientale couvertes par les établissements travaillant avec *ILS-DIAMOND*.

Au-delà du projet Mistara, la collaboration entre les partenaires du projet et l'Abes se poursuit donc !

Cécile Gobbo
Responsable du Pôle Flux et données de la Bulac
cecile.gobbo@bulac.fr

[1] L'expérience liée au projet et ses apports ont fait l'objet d'un carnet de recherche sur **https://mistara.hypotheses.org/193**

© Maxime Ruscio/BULAC

Détail d'une mistara ottomane, tablette de réglure (collections de la Bulac).

Datapoc 2.0 : un référentiel « Personnes » spécifique au Muséum national d'Histoire naturelle

Soutenu par CollEx-Persée, Datapoc 2.0 permettra d'accroître la performance des systèmes d'information pour les gestionnaires des bases de données, d'améliorer la qualité des réponses et l'accès aux données pour les utilisateurs.

Le projet Datapoc 2.0, lauréat de l'appel à projets CollEx-Persée 2019-2020, fait suite à Datapoc.mnhn.fr[1], sélectionné lors de l'appel à projets 2018-2019. L'idée a vu le jour en 2016 à l'initiative de la Direction des Bibliothèques du Muséum national d'Histoire naturelle (MNHN), avec la constitution d'un groupe réunissant des professionnels impliqués dans la construction et la gestion des systèmes d'information au Muséum. Ces systèmes très hétérogènes relèvent de chacune des missions fondamentales de l'établissement : collections naturalistes et documentaires, recherches fondamentales et appliquées, expertises institutionnelles, diffusion des connaissances et formations.

Le cœur et l'objectif final des projets Datapoc successifs reposent sur l'alignement des différents gisements et types de données que possède l'établissement (données de collecte, données de taxonomie, données d'observation, données bibliographiques, données biographiques, données archivistiques), l'hétérogénéité de ces données étant marquée autant par la nature de leur contenu que par leurs formats. La seule donnée véritablement commune à tous ces systèmes est la personne qui collecte, observe, analyse, décrit, enregistre et illustre, rédige et publie. L'alignement est donc effectué à partir de cette donnée pivot et univoque : les noms de personnes.

Le premier projet était avant tout un démonstrateur de faisabilité, réalisé à partir d'un corpus limité de noms de naturalistes contemporains ou historiques (environ 500), liés à l'établissement. Il s'agissait de tester la performance de technologies et de méthodes de traitement et d'alignement automatique de données à partir des bases disponibles, sous la forme d'un prototype d'interface orienté chercheur pour la visualisation humaine et la réutilisation par des machines des résultats et des données exposées. Après un an de développements, ponctués de tests utilisateurs, la plateforme a été publiée[2]. Considérant le vif intérêt des utilisateurs, il a été décidé de continuer l'aventure en suivant les préconisations émises en fin de projet.

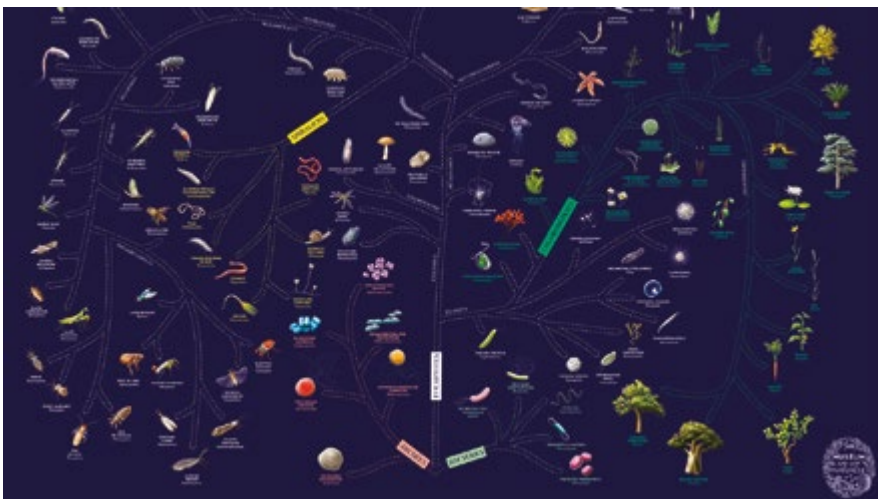
UN RÉFÉRENTIEL « PERSONNES » SPÉCIFIQUE AU MUSÉUM

Datapoc 2.0 vise donc à consolider ces premiers résultats et notamment à améliorer la qualité des données par la construction d'une base de données de référence « Personnes » propre au Muséum, sur le modèle d'IdRef. L'objectif est que chaque gisement de données au Muséum puisse utiliser ce référentiel, à la manière dont les établissements du réseau de l'Abes utilisent à ce jour le référentiel IdRef.

Pourquoi, dans ce cas, ne pas utiliser directement les systèmes d'identifiants et les bases de données d'autorités personnes existantes dans les bases de données du Muséum ? Tout simplement parce qu'elles ne se rapportent qu'aux « personnes » impliquées dans le processus de publication d'ouvrages, de revues ou de documentation. Il n'existe pas en revanche de liste d'autorités se rapportant à la fois à des personnes « publiantes » et à des personnes exerçant d'autres rôles tels que : observateurs naturalistes, collecteurs de spécimens, donateurs, etc. La richesse des personnes figurant dans les bases de données du Muséum tient dans cette extraordinaire diversité. Elle reflète la pluralité des missions de l'établissement ainsi que la pluralité des contributeurs à l'exercice de ses missions, autant dans la dimension historique de l'établissement que dans celle, plus contemporaine, des sciences participatives.

La création de ce référentiel « personnes » au Muséum permettra à la fois d'accroître la performance des systèmes d'information pour les gestionnaires des bases de données et d'améliorer la qualité des réponses et l'accès aux données pour les utilisateurs.

Grâce au soutien renouvelé du GIS CollEx-Persée, la dynamique engagée autour du projet a permis aux professionnels de l'IST, chercheurs et informaticiens de partager une même vision de ce que doivent être des données ouvertes. Pensé à l'origine comme une preuve de concept technique, le projet a surtout très vite révélé l'indispensable nécessité de porter collectivement des projets d'ampleur très structurants, puisque celui-ci trouve aujourd'hui sa place dans le projet global de refonte du système d'information du Muséum.



© MNHN/Camille Dégardin

Chloé Besombes
Chef du projet, conservatrice à la direction générale déléguée aux collections du MNHN
chloe.besombes@mnhn.fr

Cécile Callou
Responsable scientifique du projet Datapoc, maître de conférences du MNHN
cecile.callou@mnhn.fr

Cindy Lim
Chargée de mission Datapoc 2.0
cindy.lim1@mnhn.fr

[1] *PoC : Proof of concept / People of Collections.*
[2] https://datapoc.mnhn.fr

Les FID, équivalents allemands du programme CollEx-Persée, ont entrepris une coopération avec NFID, l'infrastructure nationale des données de la recherche récemment créée.

Les FID allemands : *vers une infrastructure nationale d'information*

La deuxième rencontre entre le dispositif CollEx-Persée et FID (*Fachinformationsdienste*), son équivalent en Allemagne, s'est tenue les 28 et 29 juin 2021. Ce fut l'occasion d'échanges intenses entre les représentants de ces infrastructures pour mieux connaître les développements et les expériences des uns et des autres. La question de la pérennisation s'est avérée un souci partagé en France comme en Allemagne[1].

VERS LA CRÉATION D'UNE INFRASTRUCTURE D'INFORMATION NATIONALE

En Allemagne, les nouveaux FID ont été créés comme des projets à durée déterminée pouvant bénéficier d'une prolongation, sous réserve néanmoins. Très vite, non seulement les établissements porteurs de FID mais également de nombreux chercheurs en tant qu'utilisateurs de leurs services se sont opposés à ce statut : il ne semblait pas opportun de mettre en place, à grand renfort d'engagements, des services pour la science pour les abandonner quelques années plus tard.

La *Deutsche Forschungsgemeinschaft* (DFG)[2], en réponse, a commencé par prolonger le programme de sorte que les FID puissent continuer à déposer des demandes de subvention. Puis elle a demandé aux bibliothèques de déterminer les services devant être fournis de manière permanente et ceux revêtant un caractère de projet avec un début et une fin définis, par exemple, la numérisation d'un corpus de textes. Elle a, en outre, invité l'ensemble des bibliothèques FID à s'organiser de manière à constituer une infrastructure nationale d'information qui puisse être valorisée en tant que telle auprès du public scientifique mais aussi politique. Cette auto-organisation a également pour but de créer des synergies par la coopération, par exemple au niveau technique. Par ailleurs, la DFG a commencé à étudier les possibilités de financements à long terme.

Les FID se sont donc organisés et ont rédigé un programme en 2020[3]. Ils ont créé un comité de pilotage qui coordonne un groupe de travail au niveau opérationnel. Des sous-groupes de travail ont formulé des thèmes importants qui devront être traités dans les prochaines années.

L'ARRIVÉE D'UN NOUVEL ACTEUR POUR LES DONNÉES DE LA RECHERCHE

Alors que les FID commencent à peine à fonctionner et à être perçus comme une infrastructure d'information nationale, un effort important a été effectué sur le plan politique pour créer un système national de gestion des données de la recherche. Ces dernières années ont vu la prise de conscience en Allemagne de la difficulté de créer des infrastructures d'information nationales ou coordonnées au niveau national dans un État fédéral, où la compétence en matière d'éducation et de science relève plutôt de chacun des *Länder.* En 2014, un Conseil pour les infrastructures d'information avait été instauré avec une mission consultative pour les *Länder* et l'État fédéral. Le Conseil identifia les données de recherche comme le premier grand domaine de travail urgent. Face à la numérisation croissante de la science et au mouvement de la science ouverte, des structures coordonnées sont nécessaires pour collecter, décrire, archiver et rendre accessibles les données de la recherche de manière standardisée, afin de garantir une bonne pratique scientifique et une utilisation ultérieure.

Afin de mettre en place une telle structure coordonnée, trois séries d'appels à projets ont été lancées en 2019, 2020 et 2021. Les représentants des disciplines scientifiques ont été invités à former des consortiums avec des établissements tels que des centres informatiques et des bibliothèques pour poser leur candidature. Les appels d'offres ont attiré l'attention, notamment parce que la durée prévue de cinq ans pour une première période de projet est assez longue et aussi parce qu'une pérennisation du financement a été promise. L'objectif est de créer un réseau d'environ 30 consortiums représentant les disciplines scientifiques majeures qui recevront chacun un financement annuel d'environ 2 millions d'euros. Sur sa page d'accueil, la nouvelle infrastructure formule son identité ainsi : « Dans l'infrastructure nationale des données de la recherche (NFDI), des

[1] Catherine DÉSOS-WARNIER, « Rencontres CollEx-Persée_FID : évaluer la transition, consolider les résultats, renforcer la coopération – 28 et 29 juin 2021 », Bulletin des bibliothèques de France (BBF), 30 septembre 2021 : https://bbf.enssib.fr/tour-d-horizon/rencontres-collex-persee-fid_70100

[2] La *Deutsche Forschungsgemeinschaft* (DFG) est l'organisation centrale d'autogestion de la science en Allemagne. Dotée d'un budget annuel d'environ 3,3 milliards d'euros, elle soutient des projets scientifiques de toutes sortes ainsi que des projets d'infrastructure de bibliothèques ou de centres informatiques.

[3] *Stärkung der Kooperation und Selbstorganisation im System der Fachinformationsdienste - ein Positionspapier der FID :* https://wikis.sub.uni-hamburg.de/webis/index.php/Fachinformationsdienste_für_die_Wissenschaft

données précieuses de la science et de la recherche sont systématiquement exploitées, mises en réseau et rendues utilisables durablement et qualitativement pour l'ensemble du système scientifique allemand. [...]. Les données pertinentes doivent être mises à disposition selon les principes FAIR (*Findable, Accessible, Interoperable et Reusable*) »[4].

L'INFRASTRUCTURE NATIONALE DES DONNÉES DE LA RECHERCHE ET LES FID, UNE COOPÉRATION NAISSANTE

Les objectifs et les tâches de la NFDI mentionnés ci-dessus présentent cependant des liens et des recoupements avec les FID.

Une table ronde organisée en novembre 2021, rassemblant des représentants de la DFG, des FID et des consortiums de la NFDI, ainsi que le comité directeur des FID et la direction de la NFDI, a été pour les participants l'occasion de se présenter et d'identifier des perspectives de coopération.

Quelques FID participent déjà aux consortiums NFDI issus des disciplines pour lesquelles ils offrent leurs services, mais ce n'est pas encore généralisé. Lors de cette réunion, cinq FID ont présenté les différents modes de participation à des consortiums NFDI. Étant donné que de nombreux FID exploitent de vastes systèmes d'information, la NFDI ne devrait pas créer de nouveaux instruments de recherche isolés, mais plutôt intégrer, dans la mesure du possible, des métadonnées dans les bases de données des FID. Les bibliothèques, en tant que responsables des FID, ont une longue expérience et des compétences variées dans l'élaboration des métadonnées et des fichiers d'autorité. Avec l'apparition des *Linked Open Data,* des ontologies ont été créées dans plusieurs FID. Les FID étant souvent très bien connectés à leurs communautés scientifiques, les consortiums de la NFDI ont la possibilité d'utiliser leurs réseaux pour sensibiliser les scientifiques à la question des données de la recherche.

OFFRIR UN MEILLEUR PORTEFEUILLE DE SERVICES AUX CHERCHEURS

À l'issue de la table ronde, les participants ont exprimé le souhait que les FID et les consortiums NFDI des disciplines concernées cherchent activement à établir des contacts entre eux. Ils devraient examiner les possibilités d'utiliser les services de chacune des infrastructures ou d'en tenir compte dans leurs propres développements. Comparés à cette grande initiative politique de mise en place d'une infrastructure nationale de données de recherche, les FID sont plutôt de petits acteurs. Pourtant, ils ont déjà accompli beaucoup au cours des dernières années et ont mis en place des services concrètement visibles et très utilisés. La structure nouvellement établie des FID, avec un comité de pilotage et des groupes de travail ciblés, facilite la communication avec une structure telle que NFDI. Si une coopération solide est mise en place, la science bénéficiera de la valeur ajoutée des infrastructures d'information en réseau, des données de la recherche jusqu'aux publications qui en découlent. Pour les responsables des infrastructures, la répartition des tâches et la coopération offrent la possibilité de mettre en place un meilleur portefeuille de services pour leurs communautés, plus efficace et plus économique que si des structures doubles étaient créées. Pour que cette vision devienne réalité, l'attention, l'engagement et une réflexion transversale sont nécessaires de part et d'autre. Et il est essentiel que les deux infrastructures bénéficient d'une perspective sûre de financement à long terme, car c'est la seule façon de créer des services durables pour les générations futures de scientifiques en Allemagne.

Livre de la maison de Tübingen. Livre de calendrier Iatromathématique ; l'art de l'astronomie et de la géomancie. Troisième quart du XV^e siècle.

Marianne Dörr
Universitätsbibliothek Tübingen
marianne.doerr@uni-tuebingen.de

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

[4] https://www.nfdi.de/verein/#kurzinfo (Toutes les informations sont disponibles en allemand et en anglais)

La conservation partagée des périodiques de l'enseignement supérieur : améliorer l'accès à la documentation

Grâce à la subvention de CollEx-Persée, les plans de conservation partagée, qu'accompagne le CTLes, ont pu être étendus au-delà du périmètre initial de l'Île-de-France et couvrir de nouveaux domaines.

Expérimentée avec profit depuis 2005 par les bibliothèques universitaires médicales d'Île-de-France aux espaces de stockage saturés et confrontées à l'expansion des abonnements électroniques à financer, la mutualisation de la gestion des périodiques imprimés de l'enseignement supérieur dans des plans thématiques de conservation partagée (PCP) progresse grâce au soutien ministériel et l'intermédiaire du GIS CollEx-Persée. C'est ainsi que la subvention attribuée à ce programme par le GIS est passée de 200 000 euros en 2016 à 419 000 euros en 2022. Cette augmentation, même répartie sur un grand nombre de bibliothèques, permet l'extension géographique au-delà du périmètre initial de l'Île-de-France et la couverture de nouveaux domaines, avec désormais 17 PCP dont les deux derniers ouverts en 2020, l'économie-gestion et les sciences du numérique.

L'objectif est de susciter, puis soutenir, des réseaux de partenariats thématiques entre établissements volontaires à l'occasion d'appels à projets ouverts régulièrement, pour le compte du GIS, par le Centre technique du livre de l'enseignement supérieur (CTLes), opérateur national en matière de soutien à la conservation partagée depuis 2014, en plus de sa mission première de bibliothèque régionale de dépôt en Île-de-France.

UNE GESTION RATIONNALISÉE

Chaque plan est porté par un établissement pilote scientifique. Les partenaires, avec la collaboration d'enseignants-chercheurs, définissent d'abord une liste de périodiques de niveau recherche, représentant un corpus de référence dont la conservation pérenne est jugée nécessaire. Un premier travail de fond s'engage ensuite, incontournable à cause du volume et de l'ancienneté des collections : l'amélioration du signalement catalographique (avec ajout d'ISSN, des indexations-matière, des liens de filiation, etc.) et surtout la réalisation des états de collection exacts et détaillés. Il répond directement aux objectifs du GIS d'améliorer les services aux chercheurs appuyés sur les collections, dont le signalement scientifique et précis est un prérequis à la valorisation. Il conditionne évidemment la gestion physique des collections à l'intérieur d'un PCP.

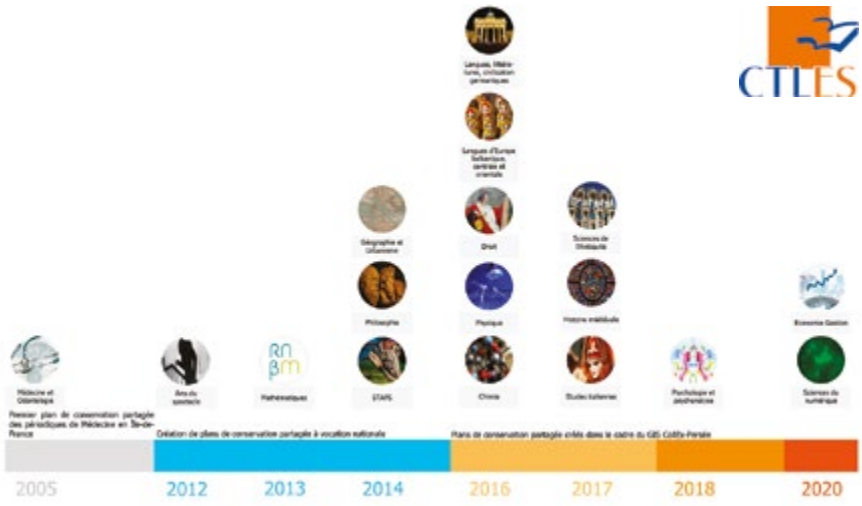
Parallèlement, un à deux établissements doivent se déclarer pôles de conservation pour chacun des titres. Cette rationalisation permet sur l'ensemble du réseau thématique un désherbage concerté qui libère de l'espace, pour de nouvelles acquisitions ou pour une reconversion totale d'un local, avec des transferts vers les pôles de conservation qui complètent ainsi leurs collections par rapport à la liste de référence prédéfinie. Des abonnements papier peuvent aussi être arrêtés, en fonction de chaque situation locale.

L'Agence bibliographique de l'enseignement supérieur (Abes), avec ses outils catalographiques nationaux et l'interface de visualisation des états de collection *Périscope,* représente un partenaire essentiel à l'avancée du programme. Le CTLes fournit, quant à lui, des outils informatiques dédiés à la gestion des PCP, ainsi qu'un support méthodologique (notamment *via* des formations) et logistique (avec une navette mensuelle pour les transferts entre établissements en Île-de-France).

Les pôles de conservation ne s'engagent pas seulement à conserver les titres mais aussi à y donner accès. La cartographie constituée doit ainsi diminuer l'éparpillement de la documentation et améliorer son signalement et son accessibilité. Elle permet d'identifier des collections redondantes candidates au programme de numérisation concertée du GIS, mis en œuvre par Persée depuis 2020, qui se développe en s'appuyant sur les réseaux collaboratifs disciplinaires établis dans le cadre du programme des PCP. À terme, une partie de cette documentation sera accessible en ligne.

Le GIS souhaite également développer la transmission électronique d'articles directement auprès du chercheur ou étudiant demandeur. Un appel à manifestation d'intérêt a été ouvert en décembre 2020. D'octobre 2021 à février 2022, les établissements gros fournisseurs rejoignent progressivement une plateforme collective de transmission spécialement développée, conforme à la règlementation qui impose toujours de sécuriser la seule transmission d'une copie papier à l'usager final dans le cas d'un document sous droits.

Guillaume Niziers
Directeur du CTLes
guillaume.niziers@ctles.fr



Dans le cadre du programme d'acquisition de ressources numériques sous licence nationale portées par le GIS CollEx-Persée, l'Abes a acquis le corpus des archives parlementaires britanniques du XIXe siècle, et choisi d'en assurer le signalement à l'aide de Calames.

Les archives du XIXe siècle *du Parlement britannique dans Calames*

Fin 2020, soutenu par un financement du GIS CollEx-Persée, l'Abes a procédé à l'acquisition sous licence nationale du corpus des archives parlementaires britanniques du XIXe siècle[1] numérisées par ProQuest et disponible via la base *Nineteenth-Century Parliamentary Papers*[2]. L'intérêt scientifique de ce corpus est considérable : regroupant plus de 80 700 documents parlementaires produits par la Chambre des Communes (*House of Commons*) entre 1801 et 1900, il constitue une source primaire incontournable pour les chercheurs en études anglaises, droit, sciences politiques ou histoire contemporaine.

LE CHOIX DE CALAMES

Une fois le corpus acquis, en coordination avec le groupe de travail CollEx-Persée en charge des acquisitions en licence nationale, l'Abes a décidé de procéder à son signalement en s'appuyant sur Calames, catalogue des archives et manuscrits de l'ESR, la structure EAD (*Encoded Archival Description*) s'avérant la plus adaptée au traitement de ce corpus. En effet, la force de Calames est de conjuguer principe archivistique de respect des fonds – décrits en arborescence grâce au format EAD – et possibilités de recherche étendue, sur chaque métadonnée, chaque document, par mots clés et par index[3].

D'ordinaire, et ce depuis près de dix ans, l'Abes utilise les formats (RDF) et technologies du Web sémantique pour modéliser et intégrer les données fournies par les éditeurs et autres fournisseurs. Pour procéder à ce traitement spécifique réalisé pour la première fois par l'Abes, il convenait donc de concevoir un processus technique permettant la conversion, l'enrichissement et la production en masse de données EAD.

DE L'INDEXATION À LA CONVERSION DES DONNÉES EN EAD

Dès les années 1980, une équipe britannique avait procédé à une première indexation du corpus, un travail titanesque, né -déjà- du constat d'un manque d'appropriation par la communauté universitaire, démunie face à une telle mine d'informations. Cette indexation originale, qui a fait l'objet d'une conversion en langage d'indexation Rameau, a largement guidé l'équipe confrontée à la nécessité de classer et ordonner ces 80 755 documents d'archives, une opération indispensable en préalable à leur traitement catalographique. Pour ce faire, l'Abes s'est appuyée sur les typologies du fonds, témoins de circuits législatifs distincts, les descriptions d'archives ayant ensuite été classées chronologiquement à l'intérieur de chaque typologie.

Afin de parvenir à la structure EAD attendue par Calames, il a ensuite fallu fusionner trois sources d'informations, soit trois lots de données livrés par le fournisseur : un copieux fichier XML au typage des données peu précis et comportant beaucoup d'informations en <note> ; des fichiers MARC ; un fichier Excel comportant la liste des documents acquis, liste faisant foi d'un point de vue contractuel. Ces trois sources ont ensuite été chargées dans une base XML. Grâce aux performances et à la fiabilité du logiciel BaseX et à la puissance du langage XQUERY, ces données hétérogènes ont pu, requête après requête, être mixées puis injectées dans la structure hiérarchique de l'EAD.

En prime, lors des opérations de conversion entre l'indexation originale et l'indexation Rameau, les correspondances ont été établies pour les concepts les plus fréquents, et regroupées dans un fichier Excel qui a complété les trois autres sources.

Ainsi, en cohérence avec le programme CollEx-Persée et conformément à sa mission de mise à disposition des ressources acquises sous licence nationale, l'Abes investit le champs du signalement des archives numérisées (non localisées dans un établissement du réseau), des archives nativement numériques et des données de la recherche en portant une attention toujours plus marquée aux besoins des chercheurs et à la valorisation de leurs travaux.

Étienne Naddeo
Responsable du réseau Calames
Département des Métadonnées et services aux réseaux de l'Abes
naddeo@abes.fr

Yann Nicolas
Responsable du Labo de l'Abes
nicolas@abes.fr

[1] https://www.licencesnationales.fr/les-corpus-acquis/archives-du-parlement-britannique-du-xixe-siecle

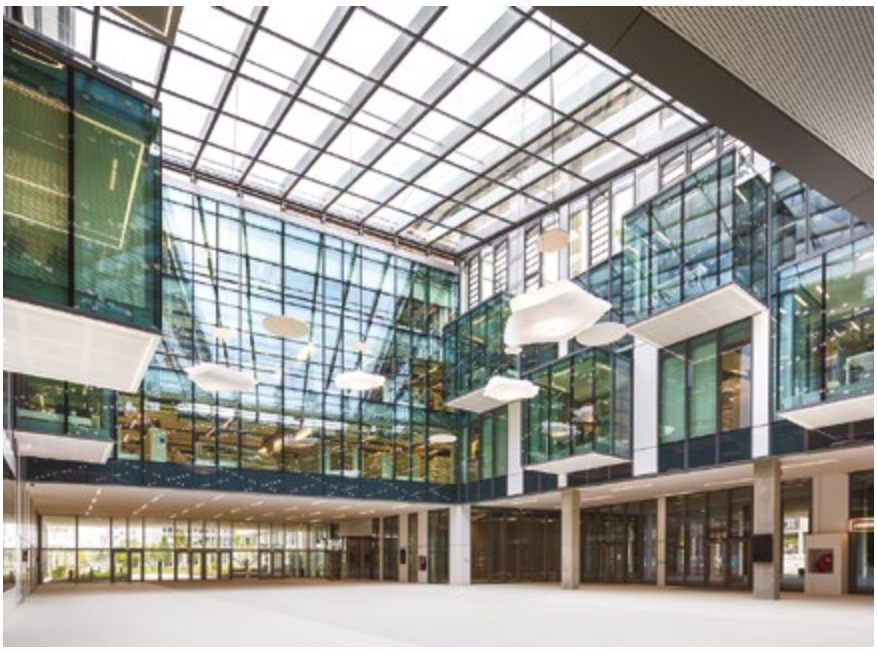
[2] https://parlipapers.proquest.com/parlipapers/search/basic/hcppbasicsearch

[3] Pour ce fonds précis, les index de lieu, de collectivité, de titre d'œuvre, de date, de provenance et de langue seront utilisés.

Bibliothèque de recherche et au service de la recherche, le Grand équipement documentaire du Campus Condorcet a été pensé comme un espace mutualisé, aussi bien en termes d'espaces, de collections que de services.

Le GED Condorcet : *un objet documentaire unique dédié aux sciences humaines et sociales*

Conçu par l'architecte franco-brésilienne Elizabeth de Portzamparc et financé par la Région Île-de-France, le Grand équipement documentaire (GED) est la toute jeune bibliothèque de recherche du tout neuf Campus Condorcet. Situé au cœur du Campus, il a vocation à en constituer le centre névralgique. La programmation du bâtiment et de ses services s'est effectuée pendant plus de 10 ans, en coconstruction étroite avec les équipes de recherche et établissements ayant vocation à s'installer sur le Campus Condorcet. Bibliothèque de recherche et au service de la recherche, le GED a été pensé comme un espace mutualisé, aussi bien en termes d'espaces, de collections que de services.



© Sergio Grazia

UN FORUM DE 1 000 M², TRAIT D'UNION ENTRE LE NORD ET LE SUD DU CAMPUS

Le bâtiment se développe sur plus de 23 000 m² et 6 étages. Il fait la part belle aux espaces professionnels qui manquaient cruellement dans les implantations de départ : 10 grandes réserves en sous-sol, protégées aussi bien d'éventuelles infiltrations que de tout passage de canalisation, grands espaces collectifs destinés au prétraitement des documents, à la numérisation, au tri des dons, bureaux organisés sur 2 étages autour d'un patio.

Côté public, on entre dans le GED par un rez-de-chaussée traversant, ouvert à toutes et tous : cette rue intérieure prolonge le Cours des Humanités, allée piétonne traversant le Campus. Elle s'élargit sous une immense verrière pour devenir un Forum de plus de 1 000 m², trait d'union entre le Nord et le Sud du Campus, lieu de discussions, de rencontres, de détente et d'échanges pour des publics variés. Un auditorium, un espace d'exposition et un *open space* destiné aux étudiant.es sont également librement accessibles au rez-de-chaussée, qui s'enrichira dans le courant 2022 d'un café et d'une librairie. L'accès aux étages se fait sur accréditation, pour consulter les collections et accéder à plus de 40 salles de travail en groupe, réservables sur place ou à distance. Certaines sont déjà utilisées par les établissements membres du Campus pour des séminaires, notamment autour de la documentation et des archives. Plusieurs balcons et une grande terrasse prolongent les espaces de travail et de lecture à l'air libre. La capacité totale est de 1400 places assises. À terme, le bâtiment permettra une ouverture H24, 7 j/7.

LES FONDS DE PLUS DE 50 BIBLIOTHÈQUES RASSEMBLÉES

Plus de 50 bibliothèques, fonds documentaires et services d'archives, historiquement dispersés sur autant de sites en Île-de-France, mutualisent leurs collections et archives. Au total 5 km linéaires d'archives et près d'un million de documents rejoignent le GED. Pour maintenir le lien étroit entre les chercheurs et la documentation, le principe retenu a été d'offrir plus de 80 % de ces collections en libre-accès : la collection documentaire se déploie sur 27 km linéaires de rayonnages, trois étages, organisés en 4 territoires de recherche thématiques (Connaissances et savoirs ; Histoire ; Textes, sens, création ; Espaces, populations, sociétés), ainsi que 4 aires culturelles (Afrique ; Amériques ; Asie ; Eurasie). Les archives et les ouvrages de la Réserve (ces derniers représentent moins de 5 % de l'ensemble des monographies) sont conservés dans les magasins en sous-sol.

La richesse de la collection documentaire provient notamment de la façon particulière dont elle a été constituée, à la fois pour et avec les chercheurs. Outre une abondante documentation imprimée publiée dans de multiples langues, les institutions ont collecté un ensemble exceptionnel de matériaux de recherche sur tous supports (tirés à part, littérature grise, travaux universitaires non publiés, documents iconographiques

et audiovisuels). L'enrichissement de ces collections est le fruit d'une collaboration entre professionnels de la documentation et chercheurs, ces derniers donnant ou léguant bien souvent les matériaux qu'ils ont collectés ou produits durant leurs travaux.

Un soin particulier est accordé aux ressources de niche destinées à des communautés de recherche restreintes représentées sur le Campus, sur des objets de recherche ou aires géographiques spécifiques, en coopération avec CollEx-Persée. Le GED est en effet bibliothèque délégataire pour la thématique Sociétés et populations (histoire sociale, démographie et sociologie).

Depuis deux ans, l'enrichissement des collections est désormais centralisé au GED et se fait en collaboration étroite avec les équipes de recherche : la mise en place de comités d'acquisition thématiques permet aux chercheuses et chercheurs de continuer à coconstruire la politique documentaire.

Autre principe fondateur, le GED associe dans une synergie étroite des collections éditées, imprimées et numériques, et des archives. Ces archives de la recherche étaient conservées dans les services archives des établissements membres du Campus, mais souvent aussi dans les centres et bibliothèques de recherche des unités, où elles côtoyaient déjà les collections documentaires. Tout dans le bâtiment et son offre de services est donc conçu pour faciliter le dialogue et la fluidité entre collections et archives : mutualisation des espaces de consultation, formations croisées des archivistes, documentalistes et bibliothécaires, développement d'interfaces de recherche communes… Les archives scientifiques, matériau inédit pour la recherche, occupent au sein de cet ensemble une place particulière : les fonds d'archives continueront à se développer par la collecte et la valorisation des archives des chercheurs et des unités installés sur le Campus. Les 5 km linéaires d'archives scientifiques du GED comprennent ainsi des fonds collectés comme matériaux pour la recherche, notamment en histoire, mais également les traces de l'activité des communautés de chercheurs et de leurs prédécesseurs. Ces archives deviendront ainsi elles-mêmes matériau de recherche et sources de données pour de nouveaux travaux, comme pour l'histoire de la recherche en SHS.

DES SERVICES À LA FOIS POUR LE SOUTIEN À LA RECHERCHE ET L'OUVERTURE SUR LE TERRITOIRE

Outil pour la recherche, le GED a comme mission d'accompagner les équipes de recherche au plus près de leurs projets et de faciliter l'innovation et l'intégration du numérique. Il bénéficie en particulier de la synergie avec Huma-Num[1] et Biblissima[2]. Des liens forts ont été construits entre le GED et les équipes de recherche par le biais d'un dispositif original, la mise en place de référent.es, qui sont des personnels du GED dont une partie du temps de travail est dédiée au relais entre les équipes et le GED : participer aux conseils de laboratoires, instruire les besoins, faire le lien entre besoins et services.

Le soutien à la recherche se développe autour de plusieurs axes prioritaires :

- **La gestion des données de recherche :** l'accompagnement des projets de recherche, en particulier sur les questions liées au traitement, à l'exposition, à la conservation et à la valorisation des données de recherche, se structure progressivement.
- **La numérisation de corpus :** les projets de numérisation de corpus se sont concrétisés grâce au soutien de CollEx-Persée. Deux projets portés par le GED en collaboration avec des équipes de recherche ont ainsi été financés, Palibr (Pensée anarchiste et libertaire en brochures) et Popp (Projet d'océrisation des recensements de la population parisienne).
- **La transcription de textes manuscrits,** collaborative et automatique, qui se développe en 2022.

Le GED a également comme mission d'accompagner les équipes et initiatives autour de la diffusion et la valorisation des savoirs, dans une démarche d'ouverture sur le territoire qui est consubstantielle au projet Condorcet dans son ensemble. Au-delà du public académique, riverains, scolaires, citoyens et curieux sont invités à fréquenter les espaces du rez-de-chaussée, qui progressivement s'animent autour d'expositions, d'événements et de rencontres en lien avec les axes de recherche des unités et coconstruites avec elles : rendez-vous autour de l'actualité éditoriale, projections de cycles de documentaires, représentations théâtrales, préparation d'une grande exposition en 2023 autour du Sport populaire. Ces espaces contribuent également à relever un des défis du Campus : comment faire connaître au grand public la recherche en sciences humaines et sociales, comment montrer ses enjeux, ses méthodes spécifiques, ses résultats ?

Stéphanie Groudiev
Directrice du Grand équipement documentaire Campus Condorcet
stephanie.groudiev@campus-condorcet.fr

[1] https://www.huma-num.fr
[2] https://portail.biblissima.fr

© Sergio Grazia

LES CHIFFRES CLÉS

- **Coût du projet : 100 millions d'euros, construction financée entièrement par la Région Île-de-France**
- **1 million de documents dont plus de 80 % en accès libre**
- **23 000 m² d'espaces de consultation, formation et partage des savoirs, répartis sur 6 étages**
- **5 km linéaires d'archives, sources issues de la recherche en sciences humaines et sociales**
- **1400 places assises et des salles de travail en groupe**
- **50 bibliothèques, fonds documentaires et services d'archives mutualisés**

KOHA, un système qui fait ses preuves

Système de gestion de bibliothèque en *open source* le plus utilisé dans le monde, Koha bénéficie d'une large communauté de partenaires travaillant à son développement.

À l'échelle mondiale, le système de gestion intégré de bibliothèque Koha[1] est installé dans tous types d'établissements (publiques, académiques, spécialisées) et est utilisé par des structures très variées : communes, universités, musées, ministères, associations.

Marshall Breeding, consultant américain indépendant et responsable de l'enquête « *The Library Technology Interoperability and Component Selection Survey* », estime, en 2017, que Koha est l'un des logiciels les plus utilisés dans le monde et le plus utilisé dans la catégorie *open-source*[2]. L'évaluation exacte du nombre d'établissements l'utilisant est complexe du fait de l'existence de différents prestataires pouvant l'installer mais aussi de sa mise en œuvre possible de façon indépendante et autonome. Toutefois, Marshall Breeding évalue à plus de 10 000 le nombre d'installations dans le monde[3].

En France, au sein du réseau Sudoc, et d'après les options des transferts réguliers[4], 37 établissements sur 152 utilisent Koha en août 2021 (soit 24,34 % du réseau). Les volumétries sont diverses mais la base de la Bulac (Bibliothèque universitaire des langues et civilisations) est la plus importante : 1 078 889 notices bibliographiques, 1 417 281 exemplaires et 101 920 adhérents dont 85 574 actifs.

L'*OPEN SOURCE* COMME GARANTIE

Un des grands avantages de Koha en tant que logiciel *open source* est qu'une large communauté de prestataires travaille à sa diffusion et à son développement dans de nombreux pays (États-Unis, Nouvelle-Zélande, Australie, Grande-Bretagne, Irlande, France, Norvège, Argentine). Tant que le logiciel est utilisé et que ses prestataires ont des clients avec des demandes fonctionnelles, le logiciel est maintenu et enrichi. Ainsi, même si un prestataire change de politique commerciale, les autres prennent le relais. L'avenir de Koha n'est pas tributaire d'éventuels rachats de société. En France, il existe trois prestataires à ce jour, ce qui offre également la possibilité de le choisir.

LE SUPPORT COMMUNAUTAIRE

La majeure partie des établissements ayant choisi Koha ont ainsi un contrat de maintenance auprès d'un prestataire qui assure le support et les mises à jour de Koha pour l'établissement. La communauté Koha, l'association KohaLa en particulier, assure le maintien d'un support communautaire complémentaire *via* entre autres une liste de diffusion, des évènements (journée d'étude, *hackathon,* formations), la diffusion d'une newsletter, le financement de nouvelles fonctionnalités plébiscitées par les adhérents. De plus, l'association propose des groupes techniques qui avancent sur des problématiques telles que la traduction, la transition bibliographique, l'amélioration des échanges d'information avec le Sudoc.

INFRASTRUCTURE TECHNIQUE ET PÉRIMÈTRE FONCTIONNEL

Koha s'installe sur des systèmes d'exploitation Linux. Il repose sur la couche technique Perl, MySQL, Plack et Memcached. Le moteur d'indexation est au choix Zebra ou Elasticsearch. Koha peut être hébergé chez le prestataire ou installé sur un serveur de l'établissement.

En dehors des modules classiques de logiciels de gestion de bibliothèques, Koha propose depuis peu un module de PEB. L'administration laisse un grand choix aux utilisateurs professionnels pour un paramétrage personnalisé. Un système de *plug-in* permet d'installer des modules complémentaires au choix des établissements.

Deux versions de Koha sont publiées par an. Des avancées spécifiques ont eu lieu ces dernières années, par exemple l'intégration avancée d'Elasticsearch, le développement de REST API pour l'interopérabilité, la pseudonymisation des données, l'ajout des nouveaux champs Unimarc IFLA.

ET L'ABES DANS TOUT ÇA ?

Depuis 2021, l'association KohaLa a décidé de travailler conjointement avec l'Abes pour envisager toutes les améliorations de Koha permettant d'avoir des flux de données de plus en plus fluides et des données encore plus fiables. Dans cette perspective, nous avons invité des membres de l'Abes à un *hackathon,* lors duquel le plug-in de contrôle des données a été proposé. Ce *plug-in* a été déposé sur le GitHub de l'Abes et présenté lors des Journées Abes 2021. KohaLa souhaite continuer de travailler avec l'Abes pour améliorer les flux et la qualité des données.

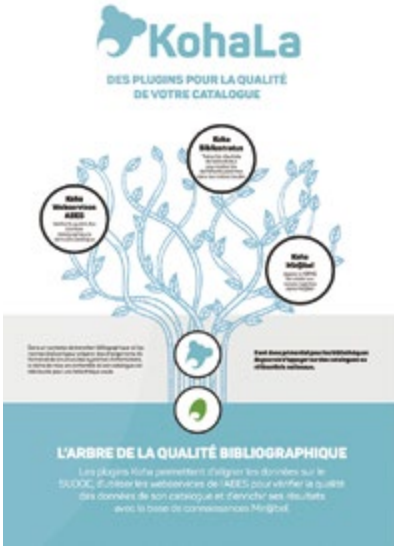
Sonia Bouis
Présidente de KohaLa
sonia.bouis@univ-lyon3.fr

[1] https://www.biblibre.com/fr/koha

[2] BREEDING Marshall, « Koha. The Original Open Source ILS », *Library Technology Reports,* 2017, vol. 53, n° 6, p. 9 17.

[3] Ibid.

[4] *Options de transferts réguliers des établissements membres du réseau Sudoc,* https://abes.fr/reseau-sudoc/reutiliser-les-donnees-sudoc/transferts-reguliers/attachment/sudoc_options_parametrage_transferts_reguliers.



Des données plus riches et mieux structurées

En janvier 2022, quatre nouvelles zones du format Unimarc bibliographique seront implémentées ou modifiées dans le Sudoc.

En début d'année, quatre nouvelles zones du format Unimarc bibliographique, conformes au format international maintenu par le *Permanent Unimarc Committee,* seront modifiées et/ou implémentées dans le Sudoc, induisant de nouvelles pratiques de signalement pour les catalogueurs. Elles permettront notamment de préciser le contexte de publication et d'utilisation de la ressource cataloguée.

- Les **identifiants de travaux universitaires,** jusqu'ici très mal documentés, pourront être enregistrés dans la zone B/029, aujourd'hui dédiée au Numéro national de thèse. Identifiants de thèses d'exercice, de mémoires de master, numéros d'ordre… aujourd'hui distribués dans d'autres zones et par conséquent mal indexés, trouveront ainsi leur place, assurant une meilleure visibilité et valorisation d'une partie de la littérature grise universitaire.

- Le signalement des reproductions numériques, lorsqu'il est associé au signalement de l'original imprimé au sein d'une même notice, n'était jusqu'à présent pas encadré dans le Sudoc. Ce sera désormais chose faite avec l'introduction de 14 nouvelles sous-zones en zone B/325, qui structureront finement les données de la reproduction numérique lorsqu'elle ne fait pas l'objet d'une notice spécifique. Les données, précédemment enregistrées en texte libre, seront ainsi plus facilement exploitables et réutilisables lors des migrations de données à venir, tout en autorisant là encore une meilleure indexation.

- Les financements, privés ou publics, nationaux ou européens, qui ont permis la publication du document catalogué pourront être signalés dans la zone B/338 dédiée à cet usage. De nouveaux index (sur le nom des projets, les sources de financement) autoriseront des requêtes fines sur ces données, leur assurant une visibilité élargie.

- La zone Unimarc B/371 enfin, dédiée aux **droits d'utilisation et de reproduction,** clarifiera notamment les droits afférents à la documentation électronique, particulièrement complexes dans l'écosystème actuel où s'entremêlent gratuité, libre accès, embargos, droits réservés et autres contraintes.

De nouvelles valeurs apparaîtront également dans les codes de fonction (zones B/7XX), le signalement des ressources continues et les codes d'écriture dans les données codées.

Laure Jestaz
Responsable du service Archives et monographies de l'Abes
jestaz@bes.fr

« BOOKSPRINT » et catalogage des ressources continues en ligne au menu de la journée des responsables des Centres du Sudoc-PS

Organisée le 14 octobre 2021 avec un délai de plus de deux ans depuis la précédente pour cause de pandémie, la réunion annuelle des responsables des Centres du Réseau Sudoc-PS a été l'occasion de retrouvailles et de découvertes plus que bienvenues.

La matinée était consacrée à une présentation des différents intervenants autour de l'application Cidemis, outil de dialogue privilégié entre l'Abes (et le Sudoc) et le Cieps (et le Registre de l'ISSN), socle bibliographique sur lequel repose le catalogage des ressources continues depuis la création du Sudoc-PS en 1983. L'après-midi a été l'occasion d'une expérimentation originale, autour de trois ateliers.

Le but : mettre à jour les consignes de catalogage pour les ressources continues en ligne, et plus particulièrement tout ce qui relève des publications en série, qu'il s'agisse de reproductions numériques de publications imprimées, de revues numériquement natives, etc.

La méthode : inspirée des « *booksprints* », elle consiste, dans un temps donné, à élaborer rapidement des contenus à partir d'une réflexion collective menée par un facilitateur (en l'occurrence un facilitateur et deux facilitatrices) et recopiée par un scribe (en l'occurrence, une scribe pour chaque atelier).

Dans l'optique LRM (*Library Reference Model*) et en en utilisant le vocabulaire, les trois ateliers ont été consacrés respectivement aux « œuvres », aux « relations entre les œuvres » et aux « items ».

L'occasion de recueillir en deux heures pour chaque atelier une riche matière, restituée aux participants sous forme de verbatims, et rédigée de façon à aboutir à la mise à jour du Guide méthodologique du Sudoc, dans le souci d'apporter aux catalogueurs réponses et consignes dans un domaine où, notamment en termes d'accès, la précision de la description bibliographique est absolument cruciale.

Si la méthode ne permet évidemment pas, en deux heures, d'aboutir à des contenus définitifs, elle est l'occasion de confronter et de conforter les points de vue d'experts, dans l'esprit de coconstruction avec ses réseaux que l'Abes met en œuvre.

Yves Desrichard
Responsable du service Ressources continues de l'Abes
desrichard@abes.fr

(Portrait)

Jean-Luc DE OCHANDIANO,

coordinateur de Prairial, le pôle éditorial Lyon-Saint-Étienne

Parlez-nous de vos fonctions actuelles…

Je suis conservateur des bibliothèques à l'université Jean Moulin Lyon 3 où j'assure la coordination de Prairial, le pôle éditorial Lyon-Saint-Étienne de soutien aux revues en accès ouvert. Nous gérons la plateforme de revues de sciences humaines et sociales du même nom et proposons un accompagnement éditorial aux revues du site Lyon-Saint-Étienne mais aussi au-delà de ce périmètre. Pour mener à bien ces missions, j'anime une équipe de sept personnes dépendant de trois établissements – les universités Lyon 2, Lyon 3 et la Maison des sciences de l'homme Lyon-Saint-Étienne – et de statuts différents – bibliothécaires, professionnels de l'édition et chargés d'information scientifique et technique. Le dispositif a donc un caractère expérimental à bien des égards mais il me paraît très pertinent car il associe des compétences complémentaires pour apporter une aide dans un domaine où les bibliothèques universitaires commencent à s'aventurer depuis quelques années.

Quelles sont les étapes qui vous semblent les plus importantes dans votre parcours professionnel ?

J'ai travaillé dans trois établissements, la bibliothèque interuniversitaire de Clermont-Ferrand, la bibliothèque de l'École centrale de Lyon et, depuis 2010, les bibliothèques universitaires de Lyon 3. Aucune expérience ne ressemble, loin s'en faut, à la suivante, ni par la taille des équipes, ni par la physionomie des collections, ni par les missions qui m'ont été confiées. C'est ce qui fait, à mon sens, la richesse du métier de bibliothécaire. Toutes ces expériences ont nourri ma réflexion sur le métier et m'ont orienté vers l'appui à la recherche qui, à son tour, m'a conduit vers le soutien à l'édition scientifique et à la science ouverte. Je parle à la première personne mais, depuis mon premier poste, j'ai appris qu'une bibliothèque est avant tout un collectif de travail et c'est en nous appuyant sur lui que nous pouvons évoluer et nous adapter.

À quand remontent vos premiers contacts avec l'Abes et dans quel contexte ?

J'ai commencé à échanger avec l'Abes quand, au milieu des années 2010, nous avons voulu généraliser l'usage des identifiants auteur au sein de Lyon 3 en nous appuyant notamment sur IdRef. J'ai alors trouvé à l'Abes des interlocuteurs attentifs à nos besoins. Mais la collaboration est devenue régulière quand le projet de diffusion d'*Arabesques* par Prairial a vu le jour. Cela a été l'occasion de mettre en place un partenariat entre l'Abes, Lyon 3, l'infrastructure de recherche Métopes et l'Inist pour créer cette version numérique de la revue et intégrer IdRef comme identifiant pivot des auteurs des articles. Là encore, les échanges ont toujours été riches et constructifs.

Quels défis majeurs, d'après vous, aura à relever l'Abes dans les prochaines années ?

L'Abes a élargi ses missions et a fortement diversifié ses activités depuis 20 ans, au point d'être aujourd'hui un acteur majeur dans le domaine de l'information scientifique et technique au sein de l'enseignement supérieur. Les défis de l'Abes sont liés à ces évolutions mêmes : innover sans perdre de vue ses missions, développer ses réseaux à l'échelle française et européenne sans oublier l'aide de proximité aux BU, notamment pour rendre possible la Transition bibliographique. Les défis sont nombreux mais l'une des forces de l'Abes est de ne jamais se couper du réseau des bibliothèques.

Qu'appréciez-vous le plus dans votre métier ?

La forte polyvalence du métier et la capacité des bibliothèques à se réinventer sans cesse en s'appuyant sur les compétences de ses agents.

Qu'est-ce qui vous énerve le plus ?

La soumission, sous toute ses formes.

Si l'Abes était un animal, ce serait… ?

Je ne pense pas qu'un seul animal suffirait à personnifier l'Abes. Je comparerais plutôt l'agence à un écosystème, au sens où l'écologie l'entend. Cet écosystème pouvait paraître assez simple au départ – une mission principale, le développement du Sudoc – mais il s'est complexifié au cours des années et il a su créer des interactions de type mutualiste avec d'autres écosystèmes proches.

Votre expression favorite ?

André Gide, dans son Journal, affirmait : « il est extrêmement rare que la montagne soit abrupte de tous côtés ». C'est un précepte important quand on veut défendre la science ouverte et la notion de biens communs.