

Proposition de feuille de route 2023 et vision à 5 ans

Conseil des membres du GIS CollEx-Persée du 13 décembre 2022

1- Rappels, contexte et démarche

L'ambition de CollEx-Persée est de **contribuer à la performance de la recherche française** en développant des services qui répondent aux nouveaux besoins des chercheurs en tant qu'utilisateurs de sources d'information et producteurs de données et de résultats de recherche. Il s'agit de mettre en œuvre tous les moyens permettant de **renforcer l'exploitation par les chercheurs**, de manière innovante, concertée et utile à l'ensemble de la communauté, des grands gisements documentaires et autres **matériaux de la recherche** constitutifs du patrimoine scientifique conservé dans les établissements de l'ESR et encore trop souvent méconnu.

- ✓ **Aux origines** : CollEx-Persée a été créé en 2017, sous la forme d'un GIS (Groupement d'Intérêt Scientifique), dans la continuité de l'infrastructure documentaire BSN (Bibliothèque Scientifique Numérique) et prenant la relève du réseau des CADIST. Elle réunit 21 établissements de recherche (membres ou associés), 4 opérateurs nationaux (Persée pour la numérisation ; l'Abes pour les données bibliographiques ; le CTLES pour la conservation des documents ; l'INIST-CNRS pour la fouille de données), la Bibliothèque nationale de France et une centaine de structures labellisées collections d'excellence. Elle constitue aujourd'hui une **plateforme de rencontres, d'échanges et de travail, unique** dans le paysage de l'ESR en ce qu'elle se concentre sur **la production et la valorisation de corpus et matériaux utiles à la recherche, de données et de services associés** dont la co-construction entre chercheurs et professionnels de la documentation garantit la pertinence, la qualité et la pérennité. Inscrite à la feuille de route nationale comme infrastructure de la recherche en information scientifique, CollEx-Persée vise à **décloisonner et dynamiser les gisements documentaires constitués afin qu'ils convergent davantage avec les attentes de la recherche aussi bien en termes de contenus que d'usages et de services.**
- ✓ **Moyens et résultats** : ses moyens d'action ont été de trois ordres : une **labellisation** CollEx destinée à donner de la visibilité aux fonds d'excellence ; des **appels à projets** corpus et nouveaux services à la recherche ou des résidences de chercheurs, destinés à rapprocher les communautés scientifiques et documentaires autour de projets communs ; des **programmes nationaux structurants**, afin de garantir la pérennité de ses résultats (autour du signalement, des licences nationales, de la numérisation ou de la préservation pérenne notamment) et financés sur la durée. CollEx-Persée n'a cessé de croître en activité depuis 2018 à travers les différents programmes transversaux ou projets (91 depuis 2018, pour un budget de 5,6 M° d'euros au total) développés au sein de son réseau. Le dispositif a été reconnu et reconduit

dans la **feuille de route nationale des infrastructures de recherche (IR)**, catégorie information scientifique (publiée le 8 mars 2022¹).

- ✓ **Penser la suite** : en 2021, un **processus d'autoévaluation** a été mené et l'année 2022 s'est ouverte avec la validation par les instances de CollEx-Persée du rapport issu de ce travail collectif². 2022 a été marquée par le renouvellement de la labellisation de CollEx-Persée comme IR, puis le lancement d'un appel à manifestation d'intérêt auprès de l'ensemble des établissements de l'ESR visant à élaborer collectivement la suite du dispositif³. Elle s'est poursuivie avec la tenue des **ateliers programmes blancs** CollEx-Persée⁴, les 11 et 12 mai 2022, qui ont réuni une centaine de participants et dont les conclusions vont alimenter la suite de la réflexion. En juin 2022, la convention constitutive du GIS étant valide jusqu'à la fin de 2023, il a été décidé de profiter de cette année pour approfondir la réflexion relative à la nouvelle feuille de route.

Le comité opérationnel et les bibliothèques associées de CollEx-Persée se sont réunis le 14 octobre, dans le but de préparer cette feuille de route et de réfléchir à la trajectoire du réseau dans les 5 prochaines années. Il a été choisi de présenter au Conseil des membres **un nombre limité de programmes suffisamment robustes**, dont le niveau de maturité fait aujourd'hui consensus au sein de la communauté et qui sont susceptibles de s'articuler autour d'une trajectoire lisible. Cette proposition est assortie **d'un axe de réflexion plus prospectif** visant à enrichir et compléter le dispositif en 2023.

La présente note a été préparée dans le cadre d'une mission de préfiguration qui a fait la synthèse des éléments recueillis dans les précédentes étapes. Elle a fait l'objet d'une concertation avec les principaux acteurs du dispositif. Elle présente **les principes directeurs, les programmes et la démarche de travail proposés au Conseil des membres pour l'élaboration de cette feuille de route**. Elle ne traite pas des aspects liés à la gouvernance du dispositif, qui seront abordés dans une autre étape et avec d'autres instances.

2- Principes directeurs de la nouvelle feuille de route

L'exercice d'autoévaluation comme les manifestations d'intérêt exprimées par les communautés et les instances scientifiques et documentaires sollicitées pour élaborer la feuille de route convergent vers **la définition de 6 principes directeurs** :

- ✓ **L'ADN de CollEx-Persée** : il se définit d'abord par **le développement des collections et des sources constitutives du patrimoine scientifique d'hier et de demain et leur valorisation pour et avec les chercheurs**. CollEx-Persée présente l'originalité d'être principalement organisé autour de processus métiers. Dans le domaine des collections, il constitue aujourd'hui l'un des rares, sinon l'unique réseau de l'ESR qui permet de conduire des initiatives ciblées décloisonnant les acteurs, les ressources, les disciplines.

¹ Actu, 8 mars 2022 : <https://www.collexpersee.eu/retrouvez-collex-persee-dans-la-nouvelle-feuille-de-route-nationale-des-infrastructures-de-recherche-2021/>

² Actu, 14 décembre 2021 : <https://www.collexpersee.eu/le-rapport-dautoevaluation-de-collex-persee-est-en-ligne/>

³ Actu, 14 février 2022 : <https://www.collexpersee.eu/pre-ami-ateliers-programmes-blancs/>

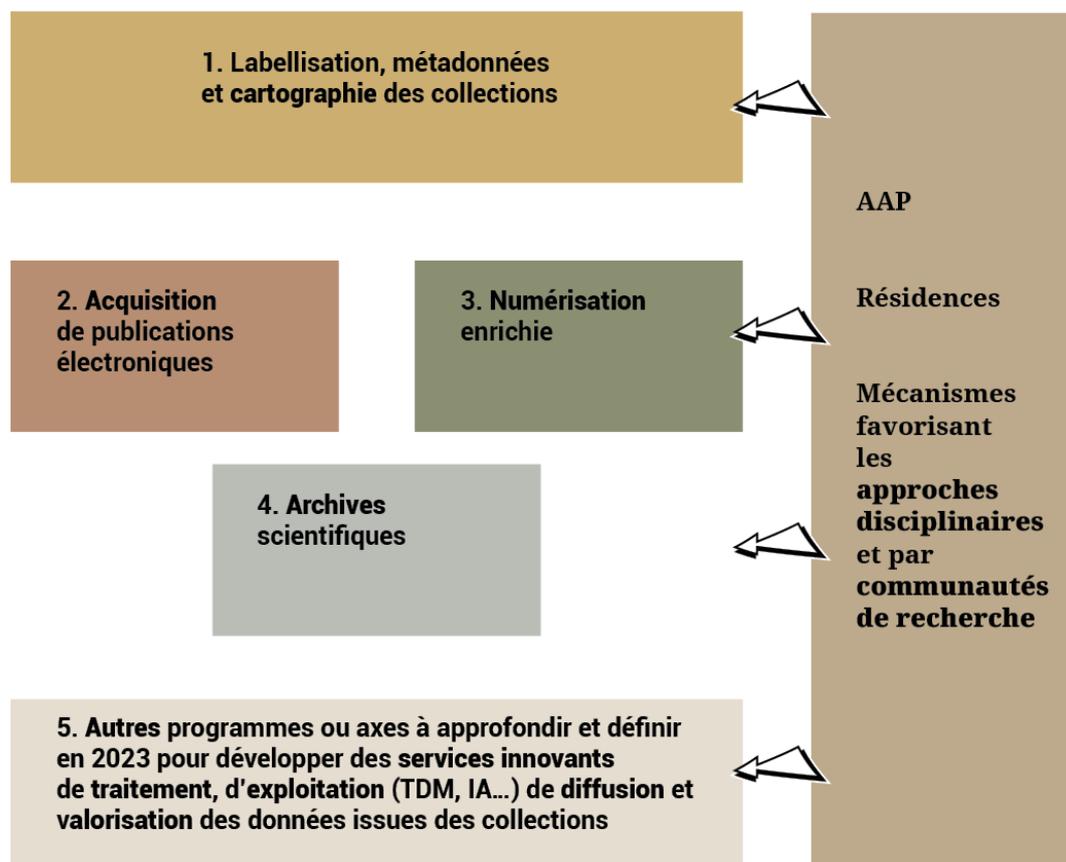
⁴ Actu, 27 avril 2022 : <https://www.collexpersee.eu/ateliers-programmes-blancs-des-11-et-12-mai-2022/>

- ✓ **Une identité lisible dans l'écosystème national** : CollEx-Persée s'inscrit naturellement dans le contexte et les priorités stratégiques de l'ESR, le plan national pour la science ouverte en particulier : les données et les outils que le dispositif produit seront conçus dans un objectif de large réutilisation, de reproductibilité et d'interopérabilité. L'infrastructure veillera également à agir en complémentarité avec d'autres IR et se concentrera sur sa valeur ajoutée. CollEx-Persée pourra enfin proposer les bonnes articulations avec les opérateurs déjà investis de missions documentaires afin d'éviter les redondances. Au niveau des établissements, son rôle est enfin celui d'un levier d'innovation et d'un accélérateur : **CollEx-Persée devrait ainsi permettre d'atteindre collectivement des résultats que les établissements seuls ne pourraient atteindre.**
- ✓ **Un dispositif de co-construction de projets entre la recherche et la documentation** : l'une des grandes forces de CollEx-Persée est sa capacité à **tisser des relations de confiance et à faciliter la conduite de projets associant chercheurs et professionnels de la documentation.** Les chercheurs sont bien placés pour exprimer leurs besoins en termes de sources comme d'outils et de services, ils apportent une expertise scientifique et disciplinaire qu'ils sont capables de mobiliser aussi bien au niveau individuel que collectif. Les bibliothécaires et professionnels de la documentation apportent un savoir-faire méthodologique, scientifique et technique qui garantit la structuration des projets et des corpus en conformité avec des normes certifiant leur interopérabilité, leur accessibilité et leur pérennité, et donc leur réutilisation et leur visibilité dans le temps long. Cette alliance est la clé du dispositif et devra être cultivée en préconisant l'association de chercheurs et de professionnels de la documentation dans la définition des corpus, le montage et le portage des programmes et projets mais aussi plus largement dans la coordination scientifique et disciplinaire du dispositif.
- ✓ **Une infrastructure de production, capable de passer à l'échelle supérieure** : la période écoulée a donné lieu à un foisonnement d'expérimentations et de projets qui ont démontré la capacité du réseau à se mobiliser autour d'une grande diversité de centres d'intérêt métier et disciplinaires. Dans un contexte compétitif et parce que la recherche doit pouvoir disposer de masses critiques de données dans des délais resserrés, il faut éviter la dispersion et **assumer une concentration des moyens autour d'un nombre limité de programmes organisés par processus métier et capables de produire et valoriser des corpus à moyenne ou grande échelle.** Fondamentalement, la mise à disposition des corpus et des connaissances associées passe par leur signalement (catalogage, inventaire, enrichissement des métadonnées...), leur dématérialisation et leur conservation en données numériques (ou bien la captation de sources nées numériques). La feuille de route pourra donc allouer une part conséquente de ses ressources à des programmes dont la finalité sera la production de données constitutives d'une collection nationale d'excellence.
- ✓ **Une infrastructure innovante dans ses domaines d'excellence** : pour autant, la culture d'innovation qui a caractérisé la période écoulée, tant dans ses approches que dans ses résultats, est un axe fort de la feuille de route. Le dispositif ne doit pas non plus être verrouillé dès sa conception mais se donner les moyens d'évoluer selon les opportunités, les innovations scientifiques et techniques. CollEx-Persée pourra ainsi utilement constituer un incubateur d'innovation capable d'héberger des projets à valeur de preuves de concept. En complément des programmes proposés à ce stade, **une réflexion sera conduite début 2023 afin d'insérer dans l'organisation du dispositif des mécanismes et des ressources favorisant le développement de services innovants pour les chercheurs.** On pense en particulier à la fouille et au traitement de données (TDM, IA...) et à la diffusion et à la valorisation des

données, thématiques qui ont donné lieu à de nombreuses propositions et manifestations d'intérêt des établissements dans l'étape de concertation.

- ✓ **Un réseau attentif à ses communautés** : afin de consolider la force du réseau qui s'est construit depuis 2017, chaque futur programme sera conçu et animé comme un "hub" ou forum de compétences, de ressources et de formation continue au sein duquel plusieurs niveaux de réseaux pourront être animés : à l'échelle d'un établissement, le dispositif sera un levier de dialogue et de collaboration entre chercheurs et professionnels de la documentation. Au niveau disciplinaire et national, ce pourra être un levier de mutualisation des contenus (corpus disciplinaires distribués) mais aussi de compétences métier (techniques, normes, cadres juridiques et réglementaires, marchés de prestation, bonnes pratiques). Des communautés d'intérêt et d'utilisateurs pourront ainsi se former autour de disciplines, d'expertises mais aussi autour d'outils. Le co-développement et la maintenance mutualisée de logiciels métiers peut être un axe fort de collaboration au sein de ces communautés. Chacun des programmes formera de fait un réseau spécialisé, le pilotage global du dispositif devant veiller à leur interconnexion au sein d'une même communauté CollEx-Persée largement ouverte à tous les professionnels et utilisateurs.

3- Propositions de programmes pour la feuille de route : synthèse



L'envergure des programmes proposés doit permettre de fédérer un grand nombre de partenaires dans et hors CollEx-Persée (autres infrastructures de recherche, opérateurs de la Culture, partenaires relevant de l'interministériel, partenaires internationaux...).

Trois d'entre eux s'inscrivent dans le prolongement des actions déjà menées dans CollEx-Persée mais avec une ampleur plus grande pour atteindre une masse critique (Cartographie d'excellence, Numérisation enrichie ou Acquisitions). **Le quatrième est novateur dans son propos autour des archives scientifiques** de l'ESR et une vaste communauté est déjà mobilisée sur ce thème. Tous s'enrichiront de projets déposés lors de l'appel à manifestation d'intérêt programmes blancs en mai 2022.

Enfin, **un champ prospectif est dégagé** autour des problématiques d'outils dont les logiciels documentaires libres utiles au signalement et à la numérisation des collections de traitement (de type IA, TDM ou autres), de valorisation ou de diffusion de données en lien avec les services à la recherche.

Les programmes sont organisés par processus métier ou par type de support et communauté d'intérêt. Les approches disciplinaires y auront néanmoins toute leur place *via* différents mécanismes destinés à alimenter les programmes, comme les appels à projets corpus et services, les appels à résidence de chercheurs, des programmes plus expérimentaux (incubateurs)...

✓ **Labellisation, métadonnées et cartographie des collections d'excellence : la tour de contrôle du réseau**

Ce programme transversal a pour mission de maintenir et développer les activités utiles à la connaissance et au pilotage global des collections labellisées, en y associant de nouvelles pratiques de signalement axées sur l'alignement et l'enrichissement de référentiels scientifiques. Il a ainsi pour objectif de mettre à disposition les outils cartographiques permettant aux communautés scientifiques de repérer les fonds d'excellence, encore trop souvent "dormants", et d'assurer la dissémination de ces données dans des projets de recherche.

Compte tenu du rôle crucial de la production de métadonnées normalisées pour garantir l'accessibilité des fonds par les chercheurs, il est indispensable de construire ce programme en association avec l'Abes et la BnF. 2023 sera l'occasion d'en définir les contours et priorités, ainsi que les moyens nécessaires en cohérence avec la stratégie des opérateurs concernés. Des partenariats pourront être recherchés avec d'autres plateformes comme data.gouv.fr ou France Archives.

A titre de pistes exploratoires, un dispositif d'AAP pourrait financer des projets collaboratifs qui exploiteraient les jeux de données produits par CollEx-Persée ; de même, le mécanisme des résidences de recherche, en complémentarité des actuelles subventions de catalogage rétrospectif de l'Abes, permettrait d'assurer la production de métadonnées enrichies sur des corpus complexes. Le sujet des données géographiques et cartographiques doit aussi s'inscrire dans ce contexte (CartoMundi).

Du fait de son rôle transverse d'appui au pilotage et à l'évaluation scientifique et documentaire, ce programme impliquera une forte coordination avec les autres, comme avec les instances scientifiques du dispositif.

✓ **Numérisation enrichie : un service de numérisation de nouvelle génération**

Dans la continuité des actions conduites depuis 2017 mais en poussant plus loin les logiques de mutualisation à la fois techniques et scientifiques, le passage à l'échelle et la qualité de production, ce programme sert trois principaux objectifs : former un centre de compétences et

une communauté métier organisés autour de la numérisation et de la production de données exploitables en veillant à l'interopérabilité et à l'ouverture des outils (cf. le soutien actuel apporté au développement de l'outil NumaHop et de sa communauté, qui doit permettre de formaliser un scénario de gestion de l'application dans une logique open source) ; financer et mettre en œuvre la numérisation enrichie de nouveaux corpus en visant des échelles significatives et un haut niveau de qualité ; assurer la préservation et la valorisation de la collection numérique constituée dans un but de capitalisation sur le long terme.

Les corpus à numériser et/ou à enrichir pourront être sélectionnés suivant différents processus déjà éprouvés et aux périmètres distincts, notamment par l'expérimentation d'une méthode de travail pour la définition concertée de priorités de numérisation de la littérature scientifique dans une approche à la fois disciplinaire et par types de documents, ou la constitution de corpus enrichis par appels à projets.

Cette approche sera complétée en s'interconnectant aux autres réflexions en cours autour de la cartographique et de la R&D (questions de transcription ou du recours à l'IA pour des opérations d'enrichissement et de fouille).

Compte tenu de ses missions et de son statut au sein du dispositif actuel, le rôle de Persée dans ce programme apparaît fondamental. Il pourrait opportunément collaborer avec un ou plusieurs autres établissements partenaires afin d'assurer le co-portage du programme.

✓ **Acquisitions de publications électroniques : la contribution de CollEx-Persée à la constitution d'un patrimoine scientifique national pérenne et de haut niveau**

Ce programme vise lui aussi à poursuivre et capitaliser les efforts déployés depuis 2017 pour l'acquisition de ressources de l'édition scientifique électronique, déjà numérisées par des tiers ou issues de l'édition nativement numérique. Ce programme est essentiel à l'ensemble de la communauté de l'ESR car il se concentre sur des supports complémentaires des achats actuels des établissements : sources, données, disciplines rares.

Il constitue par ailleurs un pool d'expertise et un centre de compétences pour les professionnels du réseau investis dans l'achat de ressources dans les disciplines et fonds couverts par CollEx-Persée, à partir du recensement périodique des besoins des chercheurs impliqués dans son réseau. Son organisation pourrait être encore davantage répartie, dans l'esprit des *Zentrum für Lizenzierung*, qui portent à la fois une dimension prospective et opérationnelle.

Il s'inscrit en complémentarité des initiatives nationales existantes que ce soit au niveau de Couperin, de l'INIST-CNRS (Istex) ou du Comité pour la science ouverte. Les contours et orientations seront précisés en 2023 en réfléchissant par exemple au rôle de levier que les financements du GIS pourraient jouer au niveau national.

L'exploitation innovante des corpus de ce programme appartient aussi au périmètre de réflexion (IA, TDM, etc.).

Le portage du programme acquisitions pourrait être assuré par un ou plusieurs établissements en liaison étroite avec les opérateurs compétents : Couperin, l'INIST-CNRS (Istex) et l'Abes.

✓ **Archives scientifiques : le forum des archives de l'ESR**

La thématique du traitement et de la valorisation des archives scientifiques a clairement émergé comme un enjeu majeur pour nombre d'établissements consultés, d'où la proposition d'y consacrer un programme dédié, centré sur ce type de support. Les archives, sous leurs formes très diverses, présentent en effet des spécificités aussi bien du point de vue de la conservation, du signalement, de la numérisation, de leur encadrement juridique que de leur utilisation

scientifique. Elles présentent aussi un fort potentiel de collaboration entre bibliothécaires, archivistes et chercheurs, dans la mesure où ces derniers sont à la fois producteurs et utilisateurs de ces fonds, au moment d'une transition numérique majeure où la production des archives des chercheurs et plus globalement des données de la recherche nécessitent de revisiter les modèles établis. C'est un domaine où les établissements de l'ESR ne sont pas toujours bien équipés alors que la demande scientifique est croissante, justifiant un investissement particulier.

Centre de compétences dédié au sujet à construire en lien étroit avec les services compétents d'autres ministères (SIAF, en particulier), ce programme pourrait d'une part être un accélérateur pour le signalement et la numérisation de fonds d'archives dans l'ESR et d'autre part offrir un cadre propice à l'exploration de nouvelles méthodes et solutions techniques pour le traitement et la conservation d'archives nativement numériques.

Plusieurs établissements ont manifesté leur intérêt pour participer au portage d'un tel programme, qui devra naturellement s'articuler avec les opérateurs de signalement (Abes) et de numérisation (Persée) de l'ESR et les programmes dédiés à ces processus.

Comme indiqué plus haut, afin de ne pas verrouiller le futur dispositif et de le doter d'un forum, d'un cadre et de ressources dédiés à l'émergence et à l'accompagnement d'autres projets, on propose que des groupes de **travail soient spécifiquement constitués en 2023 qui pourront aboutir à la constitution d'un ou deux programmes supplémentaires.**

4- Proposition de démarche pour les prochaines étapes

Fonctionnement des programmes : points d'attention

- **Coordination** : les solutions de gouvernance qui seront définies ultérieurement, devront garantir la bonne coordination des programmes entre eux et favoriser la coopération au sein du dispositif ainsi qu'avec d'autres infrastructures de recherche et initiatives au niveau national ou européen. Des projets transverses et communs à plusieurs programmes pourront être encouragés. La cohérence du dispositif passe enfin par l'emploi de standards et de technologies partagés entre les programmes.
- **Mécanismes de financement** : si le dispositif doit être resserré autour d'un nombre limité de programmes structurants pour la durée du dispositif, il laissera de la place aux initiatives et propositions des établissements, qu'il s'agisse des propositions de corpus à traiter et valoriser ou de services et d'outils à développer. Cela se concrétisera par l'élaboration concertée de mécanismes qui ont fait leurs preuves dans CollEx-Persée : guichet unique pour des appels à projets collaboratifs, résidences de chercheurs, incubateurs pour une durée courte sur des objets spécifiques (ex. des collections spécialisées comme les archives du web ou les objets de musée détenus par l'ESR) ou dispositifs plus logistiques permettant la mobilité des chercheurs et/ou des professionnels de l'information (bourses par exemple). Ces initiatives permettront d'alimenter les programmes et d'animer la dynamique de collaboration avec les chercheurs et entre les établissements qui pourront par ce canal disposer de moyens adaptés à leur participation aux programmes.

Après discussion et validation par le Conseil des membres des principes directeurs et de la proposition d'organisation des programmes, on suggère début 2023 les étapes suivantes :

- ✓ Des groupes de travail chargés d'approfondir les programmes seront mis en place. On envisage que ces groupes associent des membres des différentes instances (COMOP, CS, CM) plutôt que de conduire ces échanges de manière étanche entre les instances. L'objectif est en effet de préciser une trajectoire partagée intégrant les préoccupations professionnelles, scientifiques et stratégiques.
- ✓ Appels à manifestation d'intérêt pour le portage de chacun des programmes (le portage du programme numérisation enrichie étant assuré par Persée, qui pourra y associer d'autres établissements dans le cadre d'un AMI complémentaire);
- ✓ Les décisions relatives au portage des programmes pourront être prises en 2023 dans le cadre de la gouvernance actuelle ;
- ✓ On pourra prévoir une approche phasée dans le temps distinguant le portage des programmes les plus mûrs et ceux qui nécessiteront un travail de concertation et d'approfondissement plus conséquent ;
- ✓ Chaque programme validé pourra se voir allouer des ressources RH (responsables ou coordonnateurs de programmes) pour aider à la préfiguration et mise en œuvre opérationnelle : le BI 2023 prévoit un budget dédié qualifié de "levier à la préfiguration".