

Diffusion ouverte et préservation des collections de périodiques : pour un programme national de numérisation concertée Persée/PCP

La politique nationale de numérisation portée par le GIS CollEx-Persée prévoit un soutien à des programmes pluridisciplinaires d'envergure nationale, mobilisant différents acteurs et opérateurs du réseau. L'enjeu que représentent la diffusion ouverte et la préservation des collections de périodiques s'inscrit, par son ambition et par la diversité des acteurs concernés, dans cet axe stratégique.

Diffusant plus de 800 000 documents, Persee.fr est le portail national de diffusion des collections patrimoniales de publications scientifiques. Les choix faits par Persée de numériser des collections complètes, de proposer une structuration fine des documents et une description bibliographique au niveau de l'article et du fascicule répondent aux besoins documentaires des chercheurs. À ce titre, Persée contribue, avec les bibliothèques, le CTLe et l'Abes, à assumer les enjeux de diffusion, de signalement et de conservation des collections de périodiques ou de publications en série pour la recherche.

CollEx-Persée offre le cadre pour un programme national de numérisation concertée des périodiques, qui mobilise l'expertise de Persée en matière de numérisation et de diffusion, qui s'appuie sur la connaissance de la documentation propre à chaque domaine développée par les bibliothèques en lien avec les chercheurs, et qui puisse bénéficier du fonctionnement en réseau des plans de conservation partagée pour la gestion des collections imprimées.

Les enjeux majeurs d'un programme de numérisation concertée Persée/PCP sont :

- Optimiser la visibilité, l'accès et l'utilisation des collections de périodiques conservées par les bibliothèques ;
- Favoriser l'émergence de nouveaux usages de recherche ;
- Rationaliser la conservation des collections imprimées sans perte d'accès au contenu documentaire à travers le territoire.

1. Description du programme

Enrichissement concerté du portail Persée et diffusion des collections de périodiques des PCP

Le programme de numérisation Persée/PCP consiste à intégrer au portail Persée une sélection de titres inclus dans les PCP thématiques nationaux, en cohérence avec la politique éditoriale du portail Persée et ses orientations scientifiques, définies dans la [charte éditoriale de Persée](#) selon six axes (nature et intérêt scientifique de la publication, discipline, langue de rédaction, collection, politique de diffusion ouverte, principe de non redondance). La finalité de la numérisation est la diffusion ouverte de la publication par le biais du portail Persée.

Les plans de conservation partagée thématiques et nationaux reposent, pour chaque champ disciplinaire, sur la connaissance fine de la documentation et des besoins des chercheurs, développée par les bibliothèques partenaires. Ils offrent en outre d'importants éléments de cartographie des collections pour la recherche. Définir un programme de numérisation à partir des PCP doit permettre de donner accès à de nouveaux contenus au plus près des besoins des chercheurs : identification des segments les plus difficiles d'accès à travers le territoire, repérage des titres prioritaires selon les usages actuels constatés mais aussi selon l'importance historique dans un domaine scientifique.

Ces PCP reposent également sur un état des lieux et un signalement précis des collections conservées par les différents partenaires et sur le positionnement différencié des acteurs (en tant que pôle de conservation ou membre du plan). Associée à un programme de numérisation, cette organisation permettra :

- de garantir l'accès pour tous aux contenus documentaires, préalable à tout transfert ou désherbage de collections,
- d'identifier les collections imprimées de conservation qui constituent le complément indispensable à la numérisation opérée par Persée,
- de réunir une collection massicotable pour une numérisation plus rapide et moins onéreuse.

Accompagnement de nouveaux usages de recherche

Au-delà de la diffusion des collections, il s'agit d'accompagner l'émergence de nouveaux usages de recherche et d'exploitation des contenus numériques : fouille de texte, utilisation de référentiels, bibliométrie, etc.

Les revues numérisées dans le cadre du programme Persée/PCP

- bénéficieront des alignements de référentiels opérés pour le portail Persée ;
- alimenteront le triple store Data Persée pour une interrogation via Sparklis ou en Sparql ;
- pourront être mises en relation avec d'autres corpus de recherche, et bénéficier d'outils d'exploitation spécifiques, notamment dans le cadre des Perséides mais également à travers d'autres outils.

2. Acteurs concernés et responsabilités

(voir schéma de workflow)

Persée assure le pilotage du programme national. D'un point de vue opérationnel, Persée prend en charge (1) la numérisation ou l'intégration de documents déjà numérisés, (2) le traitement documentaire et la création de métadonnées (3) la diffusion sur le portail perseefr, l'exposition dans le web de données en format RDF dans data.persee.fr, le référencement web et la mise à disposition en OAI-PMH (4) la sauvegarde et l'export pour archivage pérenne au CINES, (5) la formation au portail Persée et à Data Persée à destination des bibliothèques.

Les **bibliothèques membres des PCP** et le **CTLes** assurent (1) l'identification des collections à numériser en priorité, en lien avec les communautés de recherche desservies, (2) l'identification d'une collection de conservation et la désignation d'une bibliothèque référente pour être l'interlocuteur privilégié de Persée, (3) la constitution et la mise à disposition d'une collection pour la numérisation et la description matérielle de la collection à numériser dans les outils de Persée, (4) la validation éditoriale et les alignements des autorités auteurs avec IdRef.

Persée assure une description fine des revues numérisées au niveau de l'article. La « validation éditoriale » est une étape finale de contrôle qualité avant diffusion. Il s'agit de vérifier les métadonnées liées aux articles (titres, auteurs, intertitres, légendes des illustrations) et d'apporter les corrections nécessaires le cas échéant. Cette tâche est effectuée dans l'outil de gestion et de production de Persée qui s'appelle jGalith. Cet outil est une application web, accessible à distance, qui sera mise à disposition des bibliothèques.

L'**ABES** garantit le signalement des collections conservées et numérisées dans le cadre des PCP.

La **BnF** met à disposition de Persée des collections déjà numérisées et présentes dans Gallica, quand la nature des fichiers permet leur intégration à la chaîne de traitement de Persée.

Le **CINES** assure l'archivage pérenne des données versées par Persée.

3. Bénéfices attendus

Pour les chercheurs

- Données bibliographiques fines, meilleur signalement/référencement de la documentation
- Accès plus facile et plus rapide aux articles et aux publications
- Usages nouveaux (outils de recherche, index, web de données, etc.)

Pour les bibliothèques

- Fiabilisation de l'accès aux collections qui favorisera les actions entreprises dans le cadre des PCP : transferts de collections vers des pôles de conservation, élimination, transfert de crédits vers l'acquisition de ressources électroniques
- Gain d'espace, optimisation de l'utilisation des locaux universitaires
- Développement de l'offre de services à destination des chercheurs
- Accompagnement des chercheurs vers le e-only
- Moyen de renforcer le dialogue avec les communautés de recherche
- Développement des compétences liées à la numérisation et aux humanités numériques

Pour CollEX-Persée

- Mise en œuvre de la politique nationale de numérisation validée en 2019
- Mobilisation de différents acteurs de CollEX-Persée et renforcement du réseau

- Action transverse à différents thèmes qui sont portés par le GIS : numérisation mais aussi PCP, conservation numérique, FDD

Pour Persée

- Planification de l'enrichissement du portail Persée (avoir une visibilité et une perspective à moyen terme)
- Appropriation de persee.fr par les bibliothèques
- Possibilité à terme de développer des sites miroirs de production

4. Calendrier

Le programme est prévu sur 3 ans, de janvier 2020 à décembre 2022. L'année 2020 constitue une phase d'expérimentation, les années 2021 et 2022 une phase de production. Une évaluation de la réalisation du programme et de son impact doit permettre de déterminer la pertinence de sa reconduction.

Phase d'expérimentation en 2020

Rencontres entre Persée et les PCP concernés

En 2019, les PCP nationaux thématiques sont au nombre de 15, dont 14 co-pilotés par le CTLES. L'année 2020 doit permettre de poser les principes d'un programme de numérisation concertée entre Persée et les PCP de disciplines entrant dans la charte éditoriale du portail. Les publications en médecine et en mathématiques ne sont pas retenues pour ce programme dans la mesure où des programmes dédiés de numérisation existent par ailleurs (Medica, Numdam). Le PCP de chimie, dont le corpus correspond aux acquisitions faites dans le cadre d'ISTEX, n'est de ce fait pas concerné. Le PCP de droit, à l'exception des publications en histoire ou en épistémologie du droit, n'est pas non plus concerné par ce programme, les périodiques de droit étant numérisés par la bibliothèque Cujas.

	En activité depuis	Piloté par
PCP Arts du spectacle	2012	BU Censier (Sorbonne-Nouvelle) + CTLES
PCP Géographie - Urbanisme	2013	BIS + CTLES
PCP Philosophie	2013	BIS + CTLES
PCP STAPS	2015	Paris-Sud + CTLES
PCP Etudes italiennes	2016	SID UGA-Grenoble INP + CTLES
PCP Europe Balkanique, Centrale et Orientale	2016	BULAC + CTLES
PCP Langues, littératures, civilisation germaniques	2016	BNU Strasbourg + CTLES
PCP Physique	2016	SID UGA-Grenoble INP + CTLES
PCP Histoire Médiévale	2017	BIS + CTLES
PCP Sciences de l'Antiquité	2017	BIS + CTLES
PCP Psychanalyse et Psychologie	2018	SCD Paris Descartes + CTLES
PCP Mathématiques	2012	RNBM
PCP Chimie	2016	SCD Lyon 1 + CTLES
PCP Droit	2016	BIU Cujas + CTLES
PCP Médecine et Odontologie	2004	BIUS + CTLES

Tester la méthode avec le PCP Antiquité

Une première coopération est engagée entre Persée et le PCP Antiquité, à partir d'une proposition de dons de la bibliothèque de l'université de Bordeaux-Montaigne.

Choix des titres prioritaires et livraison d'une première collection début 2020.

Objectifs 2020

- Établir pour les années 2021 et 2022 un programme de numérisation concertée entre Persée et les PCP volontaires
- Tester la méthode et le partage des responsabilités définis pour le programme
- Réaliser une première tranche de numérisation et traitement documentaire de 100 000 pages (mise en ligne en 2021) (voir le schéma de workflow, canal 1)
- Proposer des modules de formation à destination des bibliothèques (choix de structuration des données sur persée.fr, fonctionnalités d'interrogation, utilisation de data.persée.fr)
- Tester le fonctionnement des canaux 2 et 3 du workflow (récupération de fichiers numérisés par la BnF ou via NumaHop par les bibliothèques partenaires)

Poursuite du programme en 2021 et 2022

- Évaluation et amélioration de la méthode testée en 2020
- Réalisation des opérations de numérisation programmées avec un objectif quantitatif minimal de 100 000 pages par an. Une volumétrie plus conséquente pourra être visée en fonction du soutien financier apporté par le GIS CollEx-Persée

5. Évaluation du programme

Indicateurs pour la réalisation du programme :

- Nombre de pages numérisées dans le cadre du programme
- Nombre de documents mis en ligne sur Persée.fr dans le cadre du programme
- Nombre de titres sur lesquels les bibliothèques membres des plans se sont positionnées
- Mètres linéaires des collections qui peuvent être désherbées (collections conservées par les membres du plan pour un titre numérisé, hors collections de conservation)

Il conviendra de mesurer l'impact du programme selon des indicateurs à définir, permettant d'apprécier notamment le service rendu aux chercheurs en termes d'accessibilité de la documentation et de qualité de son signalement mais aussi en termes de renforcement du fonctionnement en réseau des structures documentaires.

6. Coût

Le coût du programme tient principalement aux ressources humaines mises à disposition par les partenaires ou recrutées spécifiquement,

1. dans le cadre des plans de conservation partagée et pour les opérations préalables à la mise en œuvre du programme de numérisation :
 - pour les opérations de constitution de corpus, récolement, signalement, transfert de collections, faisant l'objet de subventions CollEx par appel à projets triennaux
 - pour l'animation des plans par les pilotes (bibliothèques et CTLES)
2. pour la numérisation et la documentation, réalisées par Persée et faisant l'objet d'une demande de financement par CollEx-Persée pour recrutements spécifiques
3. pour la coordination du programme, la formation des partenaires et l'encadrement des contractuels, pris en charge par Persée
4. pour la participation des bibliothèques à la validation éditoriale et à l'alignement des autorités

