

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<sup>[1]</sup> 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.

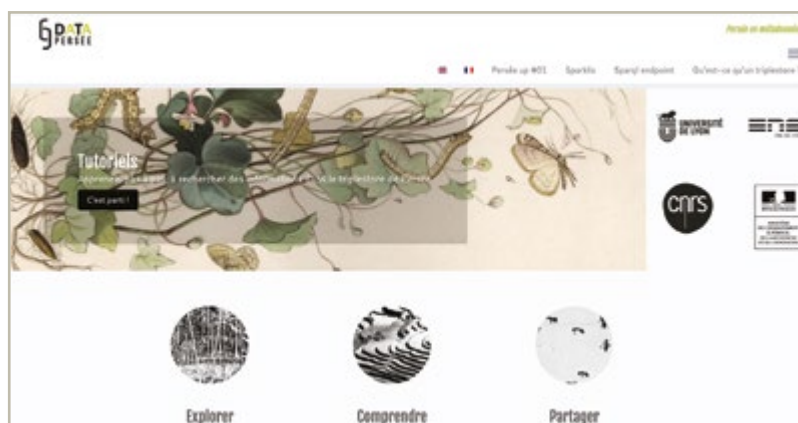
La structuration des contenus vise à diffuser les publications numérisées dans des formats qui convergent avec ceux de l'édition nativement numérique, afin d'y donner accès avec une granularité fine. L'enjeu est d'abord de faire de la numérisation des séries imprimées non seulement un vecteur de diffusion 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<sup>[2]</sup>. 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.persée.fr>

[2] <http://data.persée.fr>



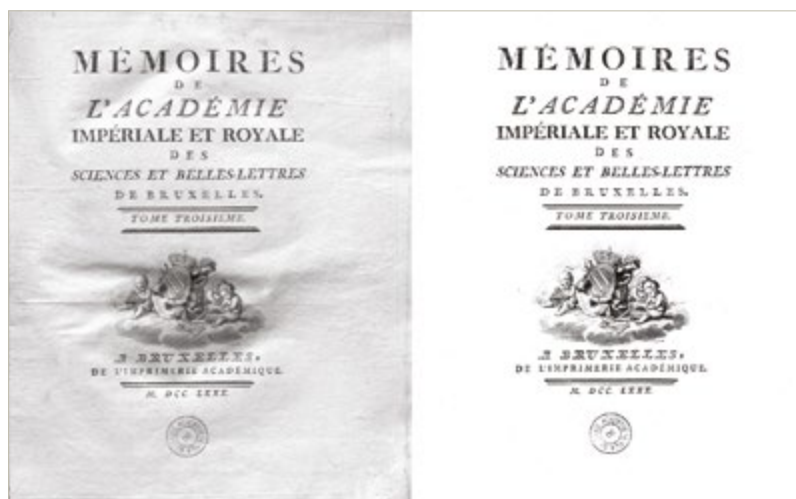
production ne peut être atteint qu'à la condition que les plateformes sur lesquelles elles sont produites, traitées et hébergées offrent les moyens non seulement de les consulter mais aussi de les manipuler. C'est dans ce but que le protocole IIIF a été implémenté sur la nouvelle plateforme de Persée 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<sup>3</sup> 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)<sup>4</sup>. 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/groupe/numerisation>

[4] Voir <https://www.collexpersee.eu/diffusion-ouverte-et-preservation-des-collections-de-periodiques-pour-un-programme-national-de-numerisation-concertee-persée-pcp>