



Valorisation numérique du patrimoine scientifique en open access. Entretien avec l'équipe Persée

Patricia Millier

Les billets d'EnssibLab
23 juin 2015

L'UMS^[1] Persée se définit comme "un service d'appui à la recherche dont les missions sont la valorisation numérique du patrimoine scientifique, la recherche et l'innovation en matière d'outils et de méthodes pour développer des corpus numériques, la diffusion de compétences et d'expertise dans ce domaine. Le portail Persée permet un accès ouvert à la connaissance : accès libre aux contenus, développement [open source](#), standards ouverts, [interopérabilité](#) et exploitation transversale des données."

L'équipe Persée, pilotée par Nathalie Fargier^[2], a bien voulu répondre à nos questions concernant les ressources en libre accès en bibliothèque et son rôle dans cet écosystème.

Pouvez-vous nous présenter le programme Persée, ses origines, ses missions ?

Persée en chiffres...

DIFFUSION

173 collections en ligne
+ 530 000 documents en texte intégral

PRODUCTION

500 000 pages / an
50 000 nouveaux documents / an

CONSULTATION (sur l'année 2014)

+ 25 millions de visites
+ 37 millions de consultations
7 millions de PDF téléchargés

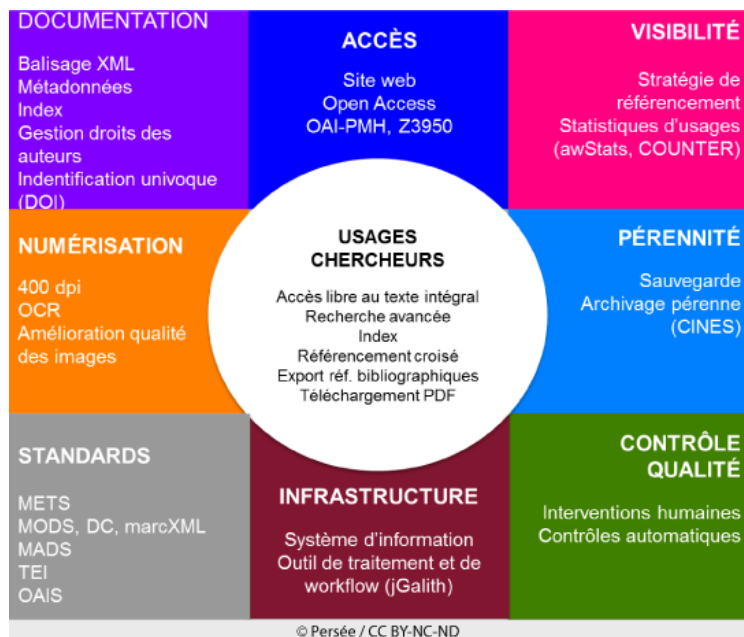
Le projet a été initié en 2003, en réponse à un appel d'offre lancé par le Ministère de l'Enseignement Supérieur et de la Recherche (MESR) dans le but d'améliorer la visibilité des résultats de la recherche scientifique francophone. Le périmètre était restreint au domaine des Sciences Humaines et Sociales, avec une typologie bien précise, celle des revues scientifiques. L'opérateur sélectionné devait remplir deux missions en parallèle : d'une part développer une chaîne de production et prendre en charge ses évolutions, d'autre part mettre en place un portail de diffusion et assurer son alimentation sur le long terme. Ce projet s'inscrivait en outre dans une dimension patrimoniale : il s'agissait de traiter l'ensemble de la revue, du premier numéro paru à la période la plus récente, en laissant à l'éditeur la possibilité de continuer la diffusion sur Persée, ou en

proposant de créer des liens vers les plateformes commerciales. Concrètement, l'ambition du projet était de proposer un accès libre et gratuit à une documentation scientifique de qualité, et d'assurer un continuum entre les contenus patrimoniaux et les sites d'édition numérique existants.

La structure retenue pour porter ce projet était initialement pilotée par l'Université Lyon 2, en partenariat avec l'ENS de Lyon, le CNRS par le biais de la Maison de l'Orient et de la Méditerranée, et l'Université de Nice - Sophia Antipolis. Après une première phase de développement des outils de production et de diffusion, une première version du portail [persee.fr](#) a été mise en ligne en 2005, avec un corpus test de 7 revues en SHS^[3], sélectionnées pour leur intérêt scientifique. La plateforme s'est depuis considérablement développée, avec une deuxième, et bientôt une troisième version du portail, et plus de 170 collections actuellement en ligne.

En 2013, Persée est entrée dans une deuxième phase en devenant une Unité Mixte de Service (UMS 3602), réunissant l'ENS Lyon, l'Université de Lyon et le CNRS, et bénéficiant toujours du soutien du MESR. Cette refonte institutionnelle a conforté Persée dans son rôle d'opérateur national tout en ancrant son action au sein du site Lyon-St Etienne. Elle a également permis de mener une réflexion de fond sur ce que devait être Persée, et de faire émerger de nouveaux axes de développement, toujours autour de cette notion fondamentale dont les mots-clés sont « la valorisation numérique du patrimoine scientifique au bénéfice de la recherche ».

En quoi vous distinguez-vous des autres bibliothèques numériques ? Ce terme vous semble-t-il d'ailleurs pertinent pour qualifier le rôle et les activités de Persée ?



Nous avons longtemps préféré le terme de "plateforme de publication électronique", sans doute dans l'idée de se démarquer des bibliothèques numériques existantes. L'ambition du projet était en effet de proposer une structuration très fine des contenus afin de satisfaire un public de chercheurs, granularité qui n'existait pas alors dans les bibliothèques numériques patrimoniales. L'unité de sens pour les chercheurs est l'article, et non le numéro ou la revue. Sous sa forme papier, un article est organisé de manière structurée (différents titres, résumé, illustrations, note de bas de page, références

bibliographiques, annexe...); nous retranscrivons cet agencement sous forme numérique pour permettre une exploitation fine du document et multiplier les accès à l'information. Ce choix de l'article comme unité documentaire de base fait sans doute la particularité de persee.fr en permettant, à partir d'une publication « papier », une articulation naturelle avec les portails diffusant les articles récents (via le protocole OAI-PMH) ou les bases bibliographiques. D'autre part, nous avons déployé des outils de navigation et de recherche équivalents à ceux proposés par les portails d'édition électronique (restitution de l'organisation éditoriale, recherche en texte intégral, export des métadonnées, téléchargement d'une version PDF, rebonds entre documents). La chaîne de production Persée a donc été pensée dans une logique assez similaire à celle de l'édition numérique native. Par exemple, ce que nous appelons le balisage XML correspond assez bien au stylage dans l'édition. Nous avons également attribué des [DOI](#) aux articles dès 2008.

Aujourd'hui, ce terme de bibliothèque numérique ne nous paraît pas du tout incohérent. Si on regarde, à un niveau très macro, les différents aspects qui peuvent caractériser une bibliothèque (ensemble d'objets organisés en collections, procédures de gestion allant de la sélection à la conservation en passant par la documentation et la diffusion), on constate que persee.fr fonctionne sur un schéma très proche (les chercheurs identifiés comme public cible, la formalisation d'une politique de développement des collections numériques). Il y a également un rapprochement à faire dans l'aspect "massif" de la bibliothèque, dans la notion de pérennité, d'ouverture et de mise à disposition des résultats de la recherche.

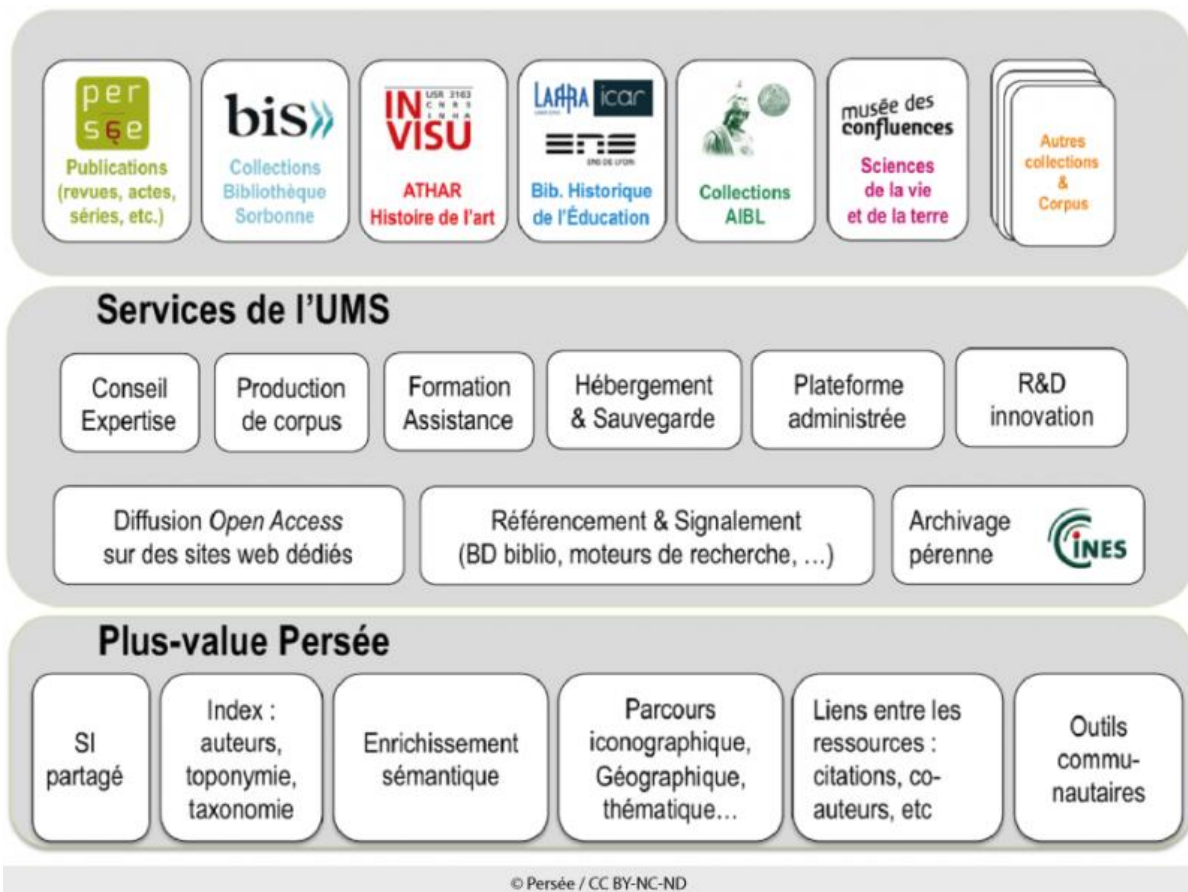
Votre modèle économique est bâti autour de l'open access et repose sur des subventions publiques (MESR et établissements de tutelle). N'est-il pas difficile de maintenir votre activité et votre développement tout en proposant un service gratuit ? Avez-vous déjà envisagé d'autres pistes de financement, comme le freemium ?

Le choix de l'*open access* est l'un des fondements du portail Persée. Tout internaute peut accéder au portail Persée, consulter les contenus et bénéficier de l'ensemble des services proposés sans payer ou sans s'identifier *a priori*. Au-delà d'une position éthique, l'ouverture et le partage sont des instruments essentiels de visibilité et de circulation de la production scientifique. Les statistiques de consultation en témoignent : 46% des visites proviennent d'autres pays que la France. Il ne s'agit pas seulement de pays francophones mais aussi des Etats-Unis, de la Chine et de pays européens non francophones. Cette politique de libre accès est permise par le fort soutien du MESR et des tutelles publiques de l'UMS. Nous n'envisageons donc aucune modification de ce modèle économique.

En revanche, pour les autres projets qui seront menés dans le cadre de l'UMS Persée, c'est la solution du co-financement qui pourrait être retenue, sur le modèle du partenariat de projet. Ce modèle n'est cependant pas complètement arrêté. Nous menons actuellement des tests pour évaluer quel pourrait être le degré d'implication de Persée en termes de ressources humaines, de temps et de financement.

Le pari de persee.fr est osé, dans la mesure où les éditeurs doivent accepter le principe d'une diffusion en libre accès du texte intégral et des métadonnées. Quels sont vos arguments pour les convaincre ?

L'intégration de nouvelles revues à persee.fr est conditionnée au respect de critères scientifiques et documentaires^[4]. Nous prenons contact avec les revues lors des manifestations comme le Salon de la Revue ou directement, mais de plus en plus, ce sont les éditeurs qui nous sollicitent. Pour les convaincre, nos arguments ont toujours été les mêmes : la visibilité offerte par le portail Persée, la valorisation des contenus (structuration fine, indexation, référencement), l'intégration de la revue au catalogue Persée, la facilité de gestion (nous prenons en charge la totalité du traitement, depuis l'acheminement jusqu'au transfert des fichiers pour archivage pérenne) et la gratuité du service (numérisation, documentation, hébergement, diffusion et archivage). Nous avons également pris en compte les demandes des éditeurs : l'autorisation de diffusion consentie à persee.fr est non exclusive, une barrière mobile ou fixe peut être définie, au-delà de laquelle les numéros jugés encore commercialisables ne seront pas diffusés. Les objections que nous rencontrions au début tenaient en fait plus à la nouveauté du programme et à la méconnaissance du numérique qu'au principe de libre accès. Lorsque les éditeurs se sont rendu compte que la diffusion sur persee.fr avait un impact favorable sur la visibilité des revues sans que cela n'affecte la vente des numéros récents, ils sont venus plus volontiers. D'ailleurs, nous avons constaté que la barrière mobile avait tendance à se réduire.



Persée entretient des rapports étroits avec d'autres acteurs de son écosystème, comme Revues.org, Cairn ou Érudit au Québec. Chaque portail signale des revues hébergées sur les autres plateformes et propose des liens vers ces collections. Comment ces partenariats sont-ils nés ?

Ils ont été envisagés dès l'origine du portail Persée. Le schéma « Erudit article » développée par *Érudit* a été implémenté par Persée, un représentant de leur équipe faisait partie du comité de suivi du projet, et nous opérons depuis un signalement réciproque de nos collections de revues. Persée.fr et revues.org ont de nombreux éléments en commun : positionnement en faveur du libre accès, valorisation de revues communes en SHS et soutien du MESR et du CNRS. La collaboration s'est très naturellement installée. *Cairn*, qui est une initiative de différentes maisons d'édition privées, propose également une offre complémentaire. Un certain nombre de collections se recoupent entre ces différentes plateformes. Tout cela a naturellement induit une logique de mutualisation.

Pourquoi avoir fait ce choix de la mutualisation plutôt que de regrouper l'ensemble des plateformes de diffusion sur un portail unique ? Est-ce une solution envisageable, envisagée ?

Ces différentes plateformes sont nées dans des contextes institutionnels différents et répondent à des logiques complémentaires. Chacune a son offre de services propre, un modèle éditorial et économique spécifique. A titre d'exemple, les revues de vulgarisation ne relèvent pas du périmètre de Persée mais relèvent du catalogue de *Cairn*. Avec *revues.org*, nous apportons une attention particulière au non recouvrement des périodes de revues que chaque plateforme diffuse. Une fusion globale risquerait de nuire à chaque projet sans pour autant améliorer la qualité de services rendus aux chercheurs. C'est pourquoi le modèle retenu est celui de la mutualisation dans le respect des périmètres déjà couverts et en veillant à une bonne articulation entre ces différents acteurs.

À l'heure de l'interopérabilité, comment gérez-vous l'échange de vos données ? Plus particulièrement, comment ces données peuvent-elles être exploitées par des structures telles que les bibliothèques ou les centres documentaires ?

Toutes les métadonnées de Persée peuvent être moissonnées via une plateforme OAI. La diversité de schémas que nous proposons, [Dublin Core](#), [MODS](#) ou [MARC XML](#), permet une large interopérabilité. Nous proposons également une passerelle Z39.50^[5] pour les bibliothèques. Nous prévoyons aussi l'ouverture d'un entrepôt de données d'ici le premier trimestre 2016. L'idée est d'aller encore plus loin dans l'"Open" : *open source* pour les logiciels, *open access* pour la diffusion, et *open data* pour les métadonnées. Actuellement, une réflexion est conduite pour associer la [Public Domain Mark](#) aux documents tombés dans le domaine public, qui sont pour l'instant sous *Creative Commons* et pour associer une licence ouverte aux métadonnées que nous produisons. Cela va dans le sens de notre positionnement en tant que plateforme de recherche : nos données doivent être libres d'utilisation.

Vous êtes en train de travailler sur une troisième version du portail Persée ; qu'est-ce qui va changer ?

La refonte du site *persees.fr* intervient dans un contexte particulier, qui est celui de la création de l'UMS Persée. Celle-ci réalisera à terme différents programmes dont *persees.fr*, qui restera le portail unique de diffusion de ce qui relève de la publication scientifique. L'enquête que nous avons menée auprès des utilisateurs de *persees.fr* entre avril et juin 2014 nous a permis de voir ce qui fonctionnait et ce qui fonctionnait moins. Nous avons repensé l'interface du site, notamment la page d'accueil qui présentera un design un peu plus éditorialisé et moderne, et proposera des informations et des actualités mises à jour régulièrement. Nous avons également fait en sorte de donner une meilleure visibilité à certaines fonctionnalités méconnues, comme l'export de références bibliographiques ou les listes d'illustrations. Nous avons d'ailleurs décidé de mettre en avant l'iconographie avec un accès facilité au niveau de l'article. De même, nous avons rendu plus visible l'accès au texte issu de l'OCR à partir de chaque page numérisée.

Côté ergonomie, le site est *responsive* ; il pourra donc plus facilement être consulté sur tablette, phablette et smartphone, même si nous restons limités par la taille minimum de lecture de l'image numérisée. Il répond également au standard du W3C concernant l'accès aux déficients

visuels.

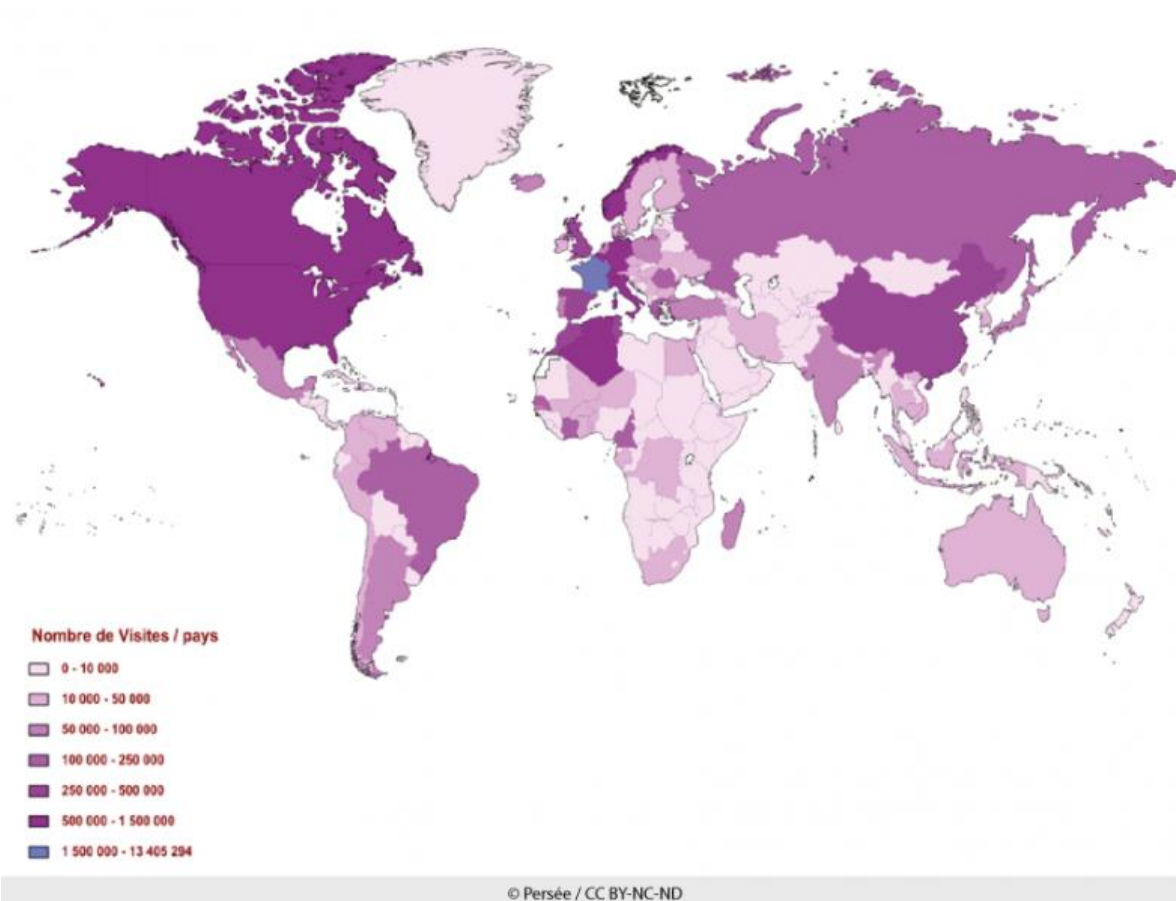
Un important travail a aussi été fait sur les fonctionnalités de recherche, que nous avons repensées en privilégiant une approche par facettes et affinement des résultats de la recherche. Nous comptons aussi mettre en place une visionneuse et proposer un widget qui permettra à l'utilisateur d'intégrer un moteur de recherche Persée sur son propre site, mais cela interviendra dans un second temps.

Le changement le plus important concerne sans doute la gestion des autorités auteurs. Dans la perspective du web de données, nous avons collaboré avec l'[ABES](#) afin d'aligner les auteurs Persée sur le référentiel IdRef. La base auteurs de Persée contient plus de 90 000 entrées, il était donc important de faire ce travail. C'est une plus-value pour le chercheur, qui aura désormais la possibilité de rebondir entre les différentes publications d'un même auteur, à l'intérieur mais aussi en-dehors du portail Persée, par exemple vers les applications de l'ABES. Cela permettra également de mettre en avant un réseau, en pointant vers les autres revues ou ouvrages dans lesquels l'auteur a écrit, ou vers les autres auteurs avec lesquels il a collaboré. Dans la même idée, des rebonds seront possibles vers Wikipedia, DBpedia, data.bnf.fr etc.

En 2014, vous avez mené une enquête de satisfaction auprès des utilisateurs du portail. Qu'avez-vous appris sur les publics de Persée ?

Nous avons lancé cette enquête^[6] dans la perspective du nouveau site, et nous avons récolté plus de 5600 réponses en 2 mois.

Sans surprise, l'enquête a révélé que nos utilisateurs étaient en majorité des étudiants en licence/master, des enseignants-chercheurs et des doctorants, ce qui correspond au public-cible du programme. Mais pas seulement ! Nous avons également des professionnels de la documentation (6%) et un certain nombre de "curieux" ce qui montre la porosité entre société et sphère scientifique. Les articles les plus fréquemment téléchargés sont certes des « classiques » des SHS mais également des articles présentant un lien fort avec des faits sociaux et politiques. Concernant la répartition géographique, il est intéressant de voir que Persée a une portée internationale : 34% des répondants venaient d'un pays étranger, principalement du Maghreb francophone et d'Afrique de l'Ouest. Nous avons d'ailleurs eu de très bons retours de ces pays, qui apprécient d'avoir accès à une documentation de qualité, qui ne leur est pas forcément proposée sur place.



Répartition mondiale des visites sur perseefr

En ce qui concerne les fonctionnalités du portail, c'est le téléchargement de PDF qui arrive en tête, notamment chez les doctorants. Étonnamment, ceux-ci connaissaient peu notre outil d'export bibliographique, ce qui nous a incités à mettre en avant cette fonctionnalité dans la nouvelle version du site.

Globalement, nos utilisateurs se sont montrés satisfaits du service, mettant en avant la qualité scientifique des contenus, leur richesse et la gratuité de l'accès. Les critiques portaient davantage sur l'ergonomie et les parcours de navigation du portail que sur les ressources proposées.

Persée, bibliothèque numérique, travaille-t-elle avec les bibliothèques "physiques" ?

Les bibliothèques sont également acteurs de perseefr. Ces partenariats peuvent prendre plusieurs formes. Le premier service rendu aux bibliothèques est la fourniture gratuite de contenus de qualité et d'une plateforme fiable de diffusion mais aussi d'archivage en lien avec le [CINES](#). Dans une perspective de conservation des collections, perseefr est un tiers de confiance et peut permettre à certaines bibliothèques de désherber leurs collections. Les bibliothèques peuvent largement moissonner les métadonnées produites par Persée et les intégrer à leur SI. Nous avons également des accords avec les principaux éditeurs d'outils de découverte qui signalent les contenus de perseefr.

Nous avons des partenariats avec le [SCD de Paris 5](#) qui dispose d'une chaîne de traitement et alimente perseefr ainsi qu'avec la [Bibliothèque interuniversitaire de la Sorbonne](#) qui utilise les outils Persée pour valoriser ses propres collections.

Nous souhaitons impliquer davantage les bibliothèques dans les futurs projets de l'UMS, en mettant en place un comité des utilisateurs où seraient intégrés des représentants de bibliothèques universitaires et des organismes de recherche, afin de réfléchir à l'enrichissement du portail, à la diversification des disciplines, en fonction des besoins des utilisateurs.

Pouvez-vous nous dire quelques mots sur les axes de développement actuels du programme ?

Ils sont intimement liés à la création de l'UMS Persée. Cette UMS réalise persees.fr, mais réalisera aussi d'autres projets. Notre plan stratégique pour 2015-2020 vise à faire de Persée un pôle de référence pour tout ce qui a trait à la valorisation numérique du patrimoine documentaire ; un lieu d'expérimentation et d'innovation en matière d'outils et de méthodes pour développer des corpus numériques ; un vecteur de diffusion de compétences et d'expertise dans ce domaine. Les missions de Persée ne se limitent plus à un champ disciplinaire (les SHS) ou à une typologie de documents (les revues).

2 activités principales sont mises en œuvre :

- continuer à enrichir le portail persees.fr en y intégrant d'une part de nouvelles disciplines – biologie, minéralogie, botanique, glaciologie, disciplines choisies pour leurs nombreuses descriptions d'espèces, qui ont encore une pertinence pour les chercheurs à l'heure actuelle ; d'autre part de nouvelles typologies – publications en séries, actes de colloques, livres.
- développer une nouvelle offre pour 2 publics cibles : les laboratoires/chercheurs et les structures de type bibliothèque, susceptibles de vouloir conduire des projets de bibliothèques numériques ou de corpus numériques. Persée pourrait alors les accompagner dans ce travail. Notre implication pourrait aller de la simple expertise jusqu'à être un véritable partenaire opérationnel, notamment en prenant en charge la numérisation et le traitement documentaire.

Nous travaillons par ailleurs à la mise en place d'un site institutionnel qui devrait être mis en ligne pour la rentrée de septembre, et qui expliquera les différentes missions de l'UMS Persée.



L'Équipe Persée / © Enssib

Notes

[1] Unité Mixte de Service.

[2] Ancienne élève de l'Institut d'études politiques (IEP) de Grenoble et de l'École nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB), Nathalie Fargier est aujourd'hui conservateur des bibliothèques et directrice de l'UMS Persée (Université de Lyon, CNRS, ENS de Lyon). La bibliothéconomie et la numérisation des collections patrimoniales sont ses domaines de spécialité.

[3] *Annales*, Bibliothèque de l'École des Chartes, *L'Homme*, *Revue de l'art*, *Revue économique*, *Revue française de science politique*, *Matériaux pour l'histoire de notre temps*.

[4] Les principaux critères sont la nature scientifique des publications, l'impact pour la communauté, l'intérêt pour la recherche en train de se faire ne se limitant pas à la seule logique patrimoniale d'histoire des sciences.

[5] Protocole de communication utilisé par les bibliothèques pour interroger simultanément plusieurs catalogues.

[6] Pour voir les résultats de cette enquête :

http://isidora.cnrs.fr/IMG/pdf/presentation_ISIDORA_Persée.pdf