

Persée, une bibliothèque numérique par et pour les chercheurs

Nathalie Fargier

▶ **To cite this version:**

Nathalie Fargier. Persée, une bibliothèque numérique par et pour les chercheurs. La francesistica italiana à l'ère du numérique, Univeristé de Gênes, SUSLLF (Società Universitaria per gli Studi di Lingua e Letteratura Francese), Institut français d'Italie, Sep 2014, Gênes, Italie. hal-01187830

HAL Id: hal-01187830
https://hal.science/hal-01187830

Submitted on 31 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Persée, une bibliothèque numérique par et pour les chercheurs

Nathalie Fargier
UMS Persée
Université de Lyon – CNRS – ENS de Lyon
nathalie.fargier@persee.fr

Résumé :
Initié en 2003 par le Ministère français de l'enseignement supérieur et de la recherche, Persée est un programme national visant à valoriser le patrimoine scientifique et à fournir aux chercheurs en sciences humaines et sociales une masse critique d'archives de publications scientifiques. Persée est devenu l'une des principales bibliothèques numériques pour les scientifiques francophones. Initialement dédié aux revues scientifiques, son périmètre s'est élargi aux publications en série, aux actes de colloques et aux livres. L'article s'attache non seulement à présenter les fondements scientifiques, politiques et technologiques de cette initiative mais aussi à exposer les interactions entre les contenus diffusés, les outils numériques de diffusion et d'exploitation et les pratiques documentaires des chercheurs. Deux aspects seront plus particulièrement mis en exergue : les services à valeur ajoutée développés dans le cadre de Persée en complément du texte intégral ; le choix d'une diffusion des contenus sans aucune restriction (politique d'*open access* préfigurant une démarche d'ouverture des données).

Mots clés : bibliothèque numérique, patrimoine scientifique, plateforme, *open access*

Introduction

Persee.fr est une bibliothèque numérique qui offre un accès entièrement libre et gratuit à des collections complètes de revues et de publications scientifiques principalement dans les domaines des sciences humaines et sociales. Il repose sur une chaîne de traitement intégrée et largement automatisée.

Initié en 2003 par le Ministère français en charge de l'Enseignement Supérieur et de la Recherche, Persée est actuellement porté par l'Université de Lyon, le CNRS et l'ENS de Lyon. Dix années après son ouverture, il continue à répondre aux besoins de son public cible et à l'accompagner dans les transformations profondes qu'une coopération quotidienne avec les directeurs de publications, les comités de rédaction et les chercheurs a su initier. Prospectif, le portail Persée s'est donné pour ambition de donner à l'exploitation et à l'archivage pérenne des documents numérisés structurés une place centrale afin de garantir l'utilisabilité et l'accessibilité à long terme du patrimoine scientifique.

Cet article présente un premier retour d'expérience et un regard réflexif sur la conception et l'évolution du portail Persée. Quels contenus sur www.persee.fr ? Quelles méthodes de valorisation ? Quel type de publication et de mise en ligne ? Quelles articulations avec la communauté scientifique ? L'objectif est de montrer de quelle manière les orientations scientifiques, les choix documentaires et les solutions techniques opérés dans le cadre Persee.fr ont permis de définir les contours d'un modèle original pour la valorisation des publications scientifiques. A l'origine pensé dans la perspective ciblée des revues en sciences humaines et sociales, Persee.fr s'est progressivement ouvert pour désormais accueillir différentes disciplines et un panel étendu de publications scientifiques (actes de colloques, livres, publications en série, etc.). Dans la continuité de cette évolution, Persee.fr est appelé à devenir une plateforme de recherche dédiée à la valorisation du patrimoine documentaire.

L'article est organisé en deux parties suivant une dialectique entre contenus et outils ; entre savoirs et appareillage technique assurant leur diffusion, leur exploitation et leur préservation à long terme. Dans un premier temps, nous nous attacherons à présenter les modalités de développement des collections de Persee.fr en mettant en lumière les critères de sélection des contenus et les procédures de structuration et de balisage de ces derniers. Le numérique permet de valoriser parallèlement la collection conçue comme un objet intellectuel à part entière et les particularités de chaque article, de multiplier le points d'accès à un ensemble structuré de documents, contextualisés, historicisés et reliés entre eux . Dans un second temps, il s'agit de présenter la plateforme Persée qui permet de traiter les collections de manière intégrée, de leur numérisation à leur archivage en passant par leur documentation, leur diffusion et leur référencement. Les objets numériques ainsi créés se distinguent fondamentalement du matériau papier de base et ils offrent des potentialités majeures en termes de recherche, de navigation et de mise en relation.

I. Construire des collections numériques patrimoniales à des fins de recherche

1. Genèse et fondements du portail Persée

En janvier 2015, nous fêterons le 10ème anniversaire de l'ouverture du portail Persée. Tout a

commencé avec sept revues [1] qualifiées de « pionnières » ce qui souligne l'aventure entreprise sur des terrains encore peu explorés en France et la prise de risque assumée. L'enjeu était tout à la fois scientifique, politique et technologique.

- *Scientifique* car il s'agissait de rendre accessible sur un site web dédié des collections complètes de revues scientifiques en sciences humaines et sociales et de rendre ce patrimoine exploitable dans une perspective de recherche ;

- *Politique* car le projet devait favoriser la visibilité de la recherche francophone au sein de l'offre scientifique internationale et le libre accès de la communauté scientifique à l'information qu'elle produit et qui est indispensable pour la poursuite de ses travaux ;

- *Technologique* enfin car l'objectif était d'assurer l'indexation des revues au niveau de l'article et de déployer des outils de navigation, d'exploitation et de recherche équivalents à ceux proposés par les portails d'édition électronique.

Cette trajectoire étant dessinée, la conception et le développement du portail Persée furent guidés par une conception fondamentalement dynamique du patrimoine et de la numérisation. Le patrimoine peut être envisagé comme un objet et un outil de la recherche ; sa numérisation est un moyen non seulement de faciliter l'accès aux résultats de la recherche française mais aussi de favoriser l'émergence de nouvelles approches épistémologiques. Il a fallu un peu plus d'un an d'échanges, d'apprentissages et de discussions entre chercheurs, bibliothécaires, éditeurs et informaticiens pour voir émerger Persee.fr comme lieu où interagissent patrimoine scientifique et technologies numériques. Persee.fr a ouvert en 2005, une première évolution majeure est intervenue en 2008 et la seconde interviendra en 2015.



Figure 1. Evolution de l'interface de Persee.fr

[1] Les Annales, Bibliothèque de l'école des Chartes, L'Homme, Matériaux pour l'histoire de notre temps, Revue de l'art, Revue économique, Revue française de science politique

A la relecture de ce calendrier, deux principales tendances émergent : la nécessité d'une performance accrue des outils et des services avec un déplacement de l'intérêt pour les documents vers une exploitation des données ; la demande d'une interaction voire d'une co-construction par les chercheurs et l'équipe qui développe Persée. Il faut donc régulièrement expérimenter de nouvelles modalités de recherche et de navigation ; développer des liens entre documents (index des auteurs, index toponymiques, référentiels) ; mettre en valeur certains éléments des documents (références bibliographiques, iconographie) ; proposer une consultation conformes aux usages communs (tablettes, smartphones) et des possibilités croissante de réappropriation (formats détachables, exports).

2. Une feuille de route pour le développement et l'enrichissement des collections

Le Portail Persée présente une forte singularité. En effet, à la différence d'autres programmes de numérisation portés par des bibliothèques, il ne dispose pas de fonds propres à valoriser mais repose sur des partenariats avec différents éditeurs scientifiques qui mettent à sa disposition des documents pour numérisation, diffusion et archivage. Persee.fr constitue et développe une collection numérique propre. Sa politique de développement est structurée autour des axes suivants :

- Réunir des documents de niveau recherche présentant un haut niveau de spécialisation ;
- Traiter des collections dans leur intégralité en mettant en valeur leur profondeur historique ;
- Sélectionner des collections présentant un intérêt pour la recherche *en train de se faire* et ne se limitent pas à la seule logique patrimoniale d'histoire des sciences *;*
- Garantir un accès universel gratuit et rapide à l'ensemble des documents.

La sélection des contenus se fonde prioritairement sur des critères scientifiques et documentaires ; des critères matériels et juridiques interviennent également mais de manière secondaire. Persee.fr diffuse des publications scientifiques c'est-à-dire des publications en direction d'un public de spécialistes, et ayant subi une forme d'examen de la rigueur de la méthode scientifique [2] . Les collections du portail Persée se composent de revues scientifiques à comité de lecture ; de comptes rendus de congrès scientifique à comité de lecture ; d'ouvrages collectifs rassemblant des articles de revue ou de recherche autour d'un thème donné, coordonnés par un ou plusieurs chercheurs appelés éditeurs ; de monographies sur un thème de recherche. La qualité intellectuelle et scientifique des documents ainsi que leur impact sur les communautés des chercheurs sont des critères déterminants. Pour évaluer ces caractéristiques, il est notamment pris en compte la présence dans les bibliothèques françaises (abonnements et politiques d'acquisition), le nombre de citations « entrantes », la reconnaissance de la collectivité éditrice et les personnalités composant le comité scientifique. Pour les revues, l'indexation dans les principales bases de données bibliographiques disciplinaires est prise en compte comme la présence dans les classements établis par l'AERES[3], la European Science Foundation et les outils de type Wos ou Scopus. Par ailleurs, la politique de soutien aux revues mise en œuvre par l'INSHS[4] constitue un autre critère permettant de prioriser l'intégration à Persée et de consolider les collections déjà diffusées. Les publications diffusées par Persee.fr sont principalement en français ; elles peuvent cependant accueillir des publications etc. dans d'autres langues. Le programme étant financé par le Ministère français de l'Enseignement Supérieur et de la Recherche et des structures universitaires françaises, la priorité est donnée aux éditeurs francophones qu'ils soient de statut public ou privé.

[2] Anthropologie, archéologie, arts, démographie, droit, économie, études classiques, études des grandes aires culturelles, études régionales, géographie, histoire, histoire des sciences et des techniques, linguistique, littérature, philosophie, psychologie, religion et théologie, sciences de l'éducation, sciences, sciences de la vie et de la terre, sciences de l'information et de la communication, sociologie.
[3] Agence gouvernementale d'évaluation des structures françaises d'enseignement supérieur et de recherche.
[4] Institut des sciences humaines et sociales du CNRS

Pour les publications dont la parution se poursuit, le Portail Persée assure une diffusion la plus complète possible, du 1er volume paru jusqu'à la période la plus récente et ce dans une logique ouverte. Les éditeurs qui souhaitent commercialiser les volumes récents sont partenaires d'autres plateformes que le Portail Persée. Dans ce cas, des liens sont établis entre le Portail Persée et ces plateformes commerciales d'édition courante.

Depuis le lancement du portail Persée, un comité scientifique dédié s'assure de la qualité des revues diffusées et de la cohérence de la collection numérique que développe Persée. Afin d'associer plus intimement de larges pans de la communauté scientifique à l'enrichissement du portail Persée, un réseau d'experts réunissant des chercheurs et des professionnels de l'information scientifique et technique va être mis en place en 2015. Ce réseau aura un double rôle : il pourra non seulement proposer d'enrichir les collections du *Portail Persée* par des publications mais aussi expertiser les demandes qui remonteraient directement auprès de Persée.

3. De la collections aux documents

Au moment de sa conception, Persee.fr fut confronté à deux ambitions : d'une part, le souhait de restituer chaque collection de revues, d'actes de colloques, de publications en série, etc. dans son intégralité avec l'évolution de son organisation propre et de ses contenus ; d'autre part, la nécessité de constituer un fonds documentaire pleinement exploitable par les chercheurs dans une logique d'homothétie avec les portails d'édition courante. Il fallait à la fois assumer une numérisation patrimoniale et une numérisation à des fins de recherche. A cette injonction *a priori* contradictoire, Persée a répondu par la juxtaposition de points de vue sur une même source, approche permise par la plasticité du numérique. Les potentialités offertes par le numérique sont utilisées pour mettre à disposition des chercheurs des documents de nature patrimoniale dans le respect de la version originelle, des métadonnées de structure et des outils de recherche et d'exploitation.

La sélection d'une collection étant intervenue, l'ensemble des documents est numérisé et océrisé et l'organisation de la collection est reconstituée virtuellement dans un modèle de données avec une grande précision. Au terme de ce processus de dématérialisation, c'est un ensemble d'objets numériques structurés fondamentalement nouveau qui est conçu. Nous partons d'un méta-objet qui dispose d'une matérialité évidente (la revue comme un ensemble des documents papier) et d'une pertinence intellectuelle (la revue en tant qu'objet éditorial) et nous opérons une déconstruction selon une approche de *Content Breakdown Structure*. Le résultat d'un tel processus est la possibilité d'une double navigation *top down* et *bottom up*, avec des allers retours entre le niveau d'agrégation le plus élevé qui est la revue et le grain le plus fin qui est l'illustration au sein d'un article.

Le portail Persée permet de visualiser les collections dans leur globalité et dans leur diversité. Les années, les tomaisons et les volumes qui organisent une revue sont restitués ainsi que les éventuels changements de titre et la structure de chaque numéro. L'identité visuelle de chaque publication est également préservée avec la diffusion du format image et du maintien des références à la page. Cependant, l'unité de sens pour les chercheurs est l'article, et non le numéro qui constitue certes le niveau pertinent de catalogage en bibliothèque mais qui présente un intérêt limité en termes de recherche. Ce choix de l'article comme unité documentaire de base fait sans doute la particularité du portail patrimonial Persée 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. Enfin, un article sous sa forme papier est organisé de manière structurée (différents titres, résumé, illustrations, note de bas de page, références bibliographiques, annexe...). Un tel agencement fait l'objet d'une retranscription sous forme numérique pour permettre une exploitation fine du document et multiplier les accès à l'information.

II. Une plateforme pour la valorisation du patrimoine scientifique

1. Une vision intégrée du traitement des documents

Le modèle mis en œuvre de Persee.fr repose sur une vision intégrée du traitement des objets documentaires : ceux-ci sont identifiés, numérisées, documentés, diffusés, référencés et archives. Au fondement du dispositif, se trouvent deux aspects complémentaires avec d'une part, le recours à des normes et des standards et d'autre part, une infrastructure matérielle et opérationnelle. La mise en œuvre de standards très largement usités au sein de la communauté garantit l'interopérabilité et l'inscription dans les circuits internationaux de circulation scientifique du portail Persée :

- OAIS pour la preservation à long terme des données et des métadonnées
- RDF[5] pour la description des données en noeuds
- METS[6] (Metadata Encoding and Transmission Standard) pour les métadonnées descriptives, d'administration et de structure
- DC[7], MODS[8], et marcXML[9] pour les métadonnées bibliographiques
- MADS[10] pour les métadonnées de type Autorités
- TEI[11] et Erudit schema pour le texte integral

Le développement Persee.fr a nécessité la conception et la mise en œuvre d'une infrastructure matérielle à même de produire, de stocker et de diffuser un volume croissant de données (plus de 4 millions de pages numérisées depuis 2005 et plus de 2 millions de visites par mois). D'un point de vue logiciel, la chaîne de traitement alimentant le portail Persée réunit un ensemble de méthodes et d'outils permettant de constituer des corpus numériques en XML en vue de leur diffusion et de leur archivage[12]. Cette chaîne fonctionne par étape séquentielle avec des contrôle qualités automatiques et humains. La numérisation constitue la première étape de cette chaîne puis intervient un ensemble de traitements (logiciel de reconnaissance optiques de caractères avec la gestion simultanée de différents alphabets, amélioration de la qualité des images, pose de marge, redressement des pages). L'étape de documentation vise à structurer chaque document, à assurer un balisage des zones documentaires et à alimenter des index. La diffusion constitue la finalité principale de la chaîne. Elle revêt un double aspect : d'abord, un site web avec une présentation de l'ensemble des collections, une éditorialisation des contenus, un outil de recherche et des parcours de navigation (www.persee.fr) puis des connecteurs qui viennent moissonner les métadonnées en vue de les intégrer à d'autres systèmes d'information (OAI-PMH et Z3950). A court terme (début 2016), il est prévu l'ouverture d'un entrepôt de données de type data.persee.fr. La valorisation d'un corpus numérique va bien au-delà de cette étape, elle passe par une véritable stratégie de référencement qui vise à signaler les contenus du portail Persée et à multiplier les accès entrant. Toute la diversité des outils de signalement est prise en compte (moteur de recherche du web, plateforme de recherche Isidore, outils de découverte, outils déployés par les bibliothèques, bases de données bibliographiques). En dernier lieu, les données primaires (images, texte issu de l'OCR) et les métadonnées apportant une plus-value informationnelle sont transférés à un tiers archiveur pour assurer leur pérennité.

[5] w3.org/RDF
[6] http://www.loc.gov/standards/mets/
[7] http://dublincore.org/
[8] http://www.loc.gov/standards/mods/
[9] http://www.loc.gov/standards/marcxml/
[10] http://www.loc.gov/standards/mads/
[11] http://www.tei-c.org/index.xml
[12] L'application conçue et développée en interne s'appelle jGalith : http://jgalith.persee.fr/jGalith/

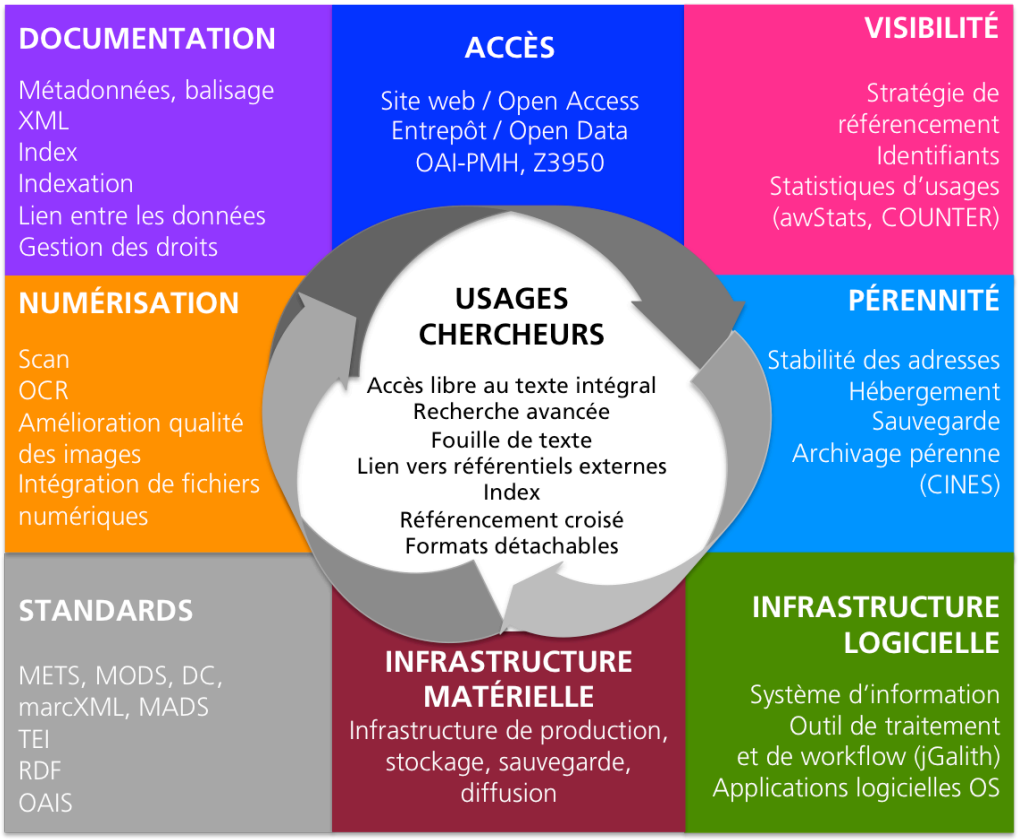


Figure 2 – Schématisation de la plateforme Persée

2. *Open Source, Open Access, Open Data*

Dès la phase de conception du portail Persée, la voie de l'*Open Access* a été retenue pour la diffusion sans aucune restriction des métadonnées et des documents en texte intégral. Tout internaute, quel que soit sa nationalité et son statut professionnel peut accéder au portail Persée, consulter les contenus et bénéficier des services proposés sans conditions de financement ou d'identification *a priori*. Cette orientation initiale – permise par le fort soutien du Ministère français de l'Enseignement Supérieur et de la Recherche et des tutelles publiques - a irrigué l'ensemble du projet. Au-delà d'une position politique, l'ouverture et le partage sont considérés comme des instruments essentiels de visibilité et de circulation de la production scientifique dans un environnement où la langue française n'est plus en situation d'hégémonie. Selon une suite logique, les développements informatiques ont été opérés dans un esprit *Open Source* avec le choix de la licence GPL. Actuellement, une réflexion est conduite pour parachever cette démarche en associant des licences ouvertes aux métadonnées et aux documents. En complémentarité de la licence Creative Commons pour les documents couverts par le droit de la propriété intellectuelle, il est envisagé de diffuser les documents libres de droit avec la licence Public Domain Mark et de diffuser les

métadonnées sous licence ouverte[13]. Enfin, avec l'ouverture d'un entrepôt de données, l'enjeu est non seulement d'utiliser les technologies du web sémantique mais aussi de mettre à disposition l'ensemble des données produites dans le cadre du portail Persée.

3. Exploitation des documents, exploitation de données

Il ne s'agit pas ici de dresser la liste exhaustive des services et des outils proposés par le biais de Persée.fr. L'objectif est davantage de présenter une typologie organisée en trois branches permettant de comprendre la démarche qui fut la nôtre et la tentative de se rapprocher au mieux des pratiques informationnelles des chercheurs. L'accès au texte intégral sans aucune restriction constitue le service de base rendu par le portail Persée. A cela s'ajoutent un outil de recherche, des fonctionnalités de navigation transverse et des possibilités de réappropriation des documents.

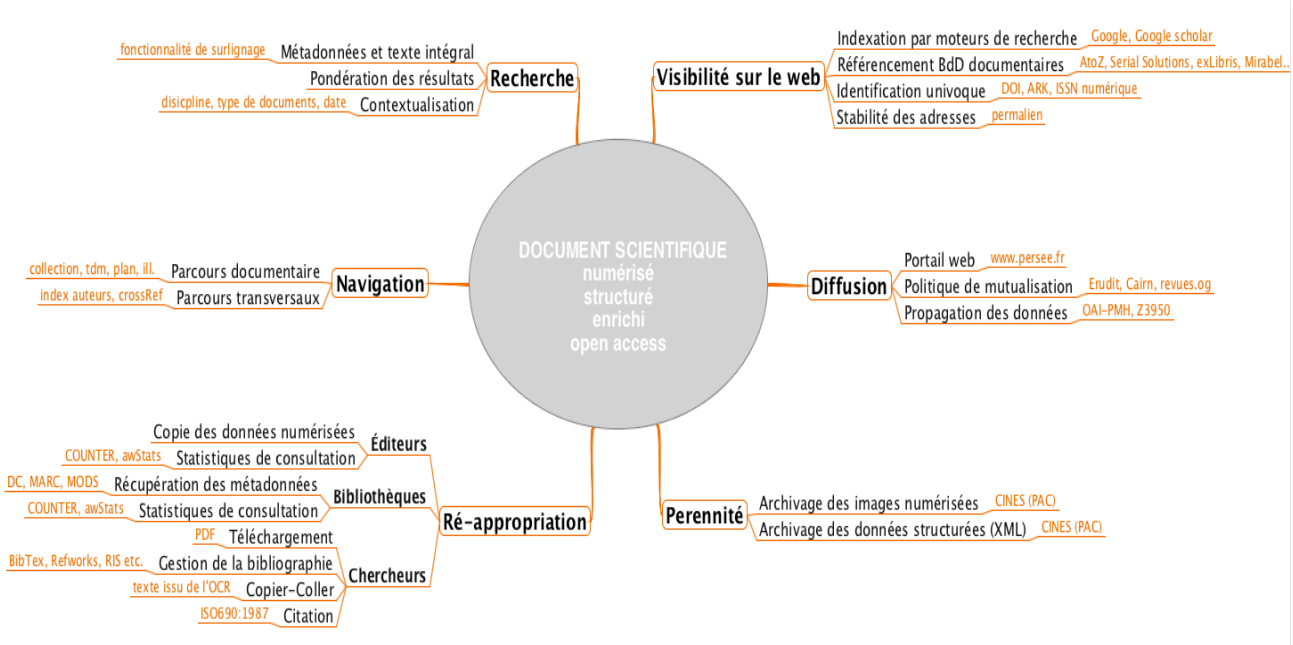


Figure 3 - Carte heuristique de l'offre de services du portail Persée

Recherche

La recherche constitue bien évidemment un enjeu majeur alors que la masse des documents diffusés par le portail Persée atteint un seuil critique. Se posent les questions de périmètre de la recherche et de performance de l'outil utilisé. Le principe retenu est celui d'une recherche dans les métadonnées et le texte intégral avec une pondération des résultats en fonction du positionnement du terme recherché. Le document contenant les mots clés recherchés dans ses métadonnées principales (titre, résumé, intertitres) et avec une fréquence importante dans le corpus du texte sera signalé parmi les premiers résultats remontés. Par ailleurs, le choix a été fait d'une recherche contextualisée. Le périmètre des documents est automatiquement redimensionné en fonction des choix de navigation de l'utilisateur : il pourra interroger l'ensemble des documents du portail, l'intégralité des articles d'une revue ou il pourra rechercher un terme dans un article, ces derniers seront alors surlignés pour faciliter leur repérage. Dans l'option la plus large, un certain nombre de critères permettra à l'utilisateur d'affiner progressivement sa recherche en combinant différentes facettes (choix d'un type de document, d'une ou de plusieurs revues, dates, etc.).

Navigation

Nous avons précédemment décrit la logique de description d'une collection que nous avons suivie et qui a permis de proposer sur Persee.fr une navigation arborescente permettant de

[13] www.etalab.gouv.fr

sélectionner une revue, une année, un volume, un numéro puis un article et de feuilleter ce dernier en suivant un plan et une liste d'illustrations. Cette représentation de l'organisation éditoriale d'une revue permet de trouver un article recherché à partir d'une référence bibliographique. Est-ce l'aboutissement de la recherche documentaire d'un chercheur ? Il est souvent pertinent de connaître d'autres articles du même auteur, de consulter parallèlement différents comptes rendus d'un même livre et de savoir quels articles citent l'article en cours de consultation et les articles auxquels ce dernier fait référence. Avec une telle contextualisation, l'article n'est plus une finalité en soi mais un nœud d'un maillage riche qui se fonde sur des données présentes en son sein. Le portail Persée met d'ores et déjà en œuvre un mécanisme de référencement croisé : des liens existent entre les articles diffusés par Persée et des plateformes extérieures peuvent pointer vers des articles de Persée auxquels nous avons attribué un DOI.

Pour parcourir plus avant cette voie particulièrement féconde, la nouvelle version du portail Persée sera dotée d'un index des auteurs avec un alignement sur des référentiels nationaux et internationaux (idRef, ISNI et VIAF) et des liens vers DBpedia. Une expérimentation est également à l'œuvre pour développer un outil permettant l'indexation de toponymes dont le déploiement bénéficiera tout particulièrement aux archéologues et aux historiens. L'expérimentation en cours est conduite à partir des « Procès-verbaux et rapports du Comité de conservation des monuments de l'art arabe », parus au Caire et en français, de 1882 à 1953, ainsi que d'un index général publié en 1914. Il s'agit d'un inventaire unique des monuments islamiques et coptes du Caire et des chantiers de restauration. Le projet est porté en partenariat avec le laboratoire inVisu[14] et l'IFAO[15]. L'objectif est de regrouper les variantes du nom d'un monument autour d'une forme de référence, afin de pouvoir rassembler tous les documents se rapportant à un même édifice, quelle que soit la graphie de son nom et les variantes de traduction et de translittération. Les outils de Persée sont utilisés pour repérer des entités nommées au sein d'un texte et un module est en cours de développement pour établir des liens avec des référentiels extérieurs. Sur Persee.fr, un nuage de termes réunissant la forme de référence et toutes les formes liées permettra d'accéder à toutes les occurrences existant et présentées dans leur contexte. Le format SKOS sera utilisé pour ce référentiel.

Réappropriation

Enfin, le portail Persée permet aux chercheurs de se réapproprier les métadonnées et les contenus diffusés. Les références bibliographiques peuvent être exportés dans la plupart des outils de gestion bibliographique et elles sont également disponibles dans Zotero. Pour les articles, les chercheurs récupèrent un fichier PDF enrichi qui réunit non seulement le fac-similé de l'article papier pour disposer d'une version conforme à l'original et le texte issu de l'OCR pour effectuer des recherches dans le texte intégral mais aussi l'ensemble des métadonnées liées à l'article (référence bibliographique, plan, liste des illustrations, résumé et citations le cas échéant).

Conclusion

Si le tournant du numérique impose des contraintes techniques, il induit également une remise en cause profonde des modalités de production et de diffusion des résultats de la recherche ainsi qu'une démarche par hybridation des compétences et des métiers. En France, Persée a participé à cette évolution fondamentale en impliquant les chercheurs en sciences humaines et sociales et les éditeurs scientifiques dans un projet ambitieux de numérisation, de valorisation et d'enrichissement du patrimoine scientifique. La réflexion partagée et le fonctionnement par-delà les périmètres habituels des métiers ont permis de

[14] http://invisu.inha.fr/

[15] http://www.ifao.egnet.net/

fonder le portail Persée au plus près des besoins de la communauté scientifique et de préfigurer un rapprochement entre numérisation patrimoniale et édition électronique. L'évolution du portail Persée témoigne d'un besoin croissant de performance des outils et de fouille des documents pour atteindre la donnée. Suivant cette tendance, le portail Persée est en train d'évoluer fondamentalement vers une organisation de ses données sous forme de graphes et une plus grande intégration dans le réseau global des données structurées.

Annexe 1 : Persee.fr en chiffres

Nombre de documents en ligne, en texte intégral	496 173
Nombre de collections en ligne (revues, publications en série, collection de livres, etc.)	157
Nombre de visites (année 2014)	25 155 417
Nombre de téléchargements PDF (année 2014)	6 932 788
Volume de données archivées	14 To

Annexe 2 : Outils et services de Persee.fr

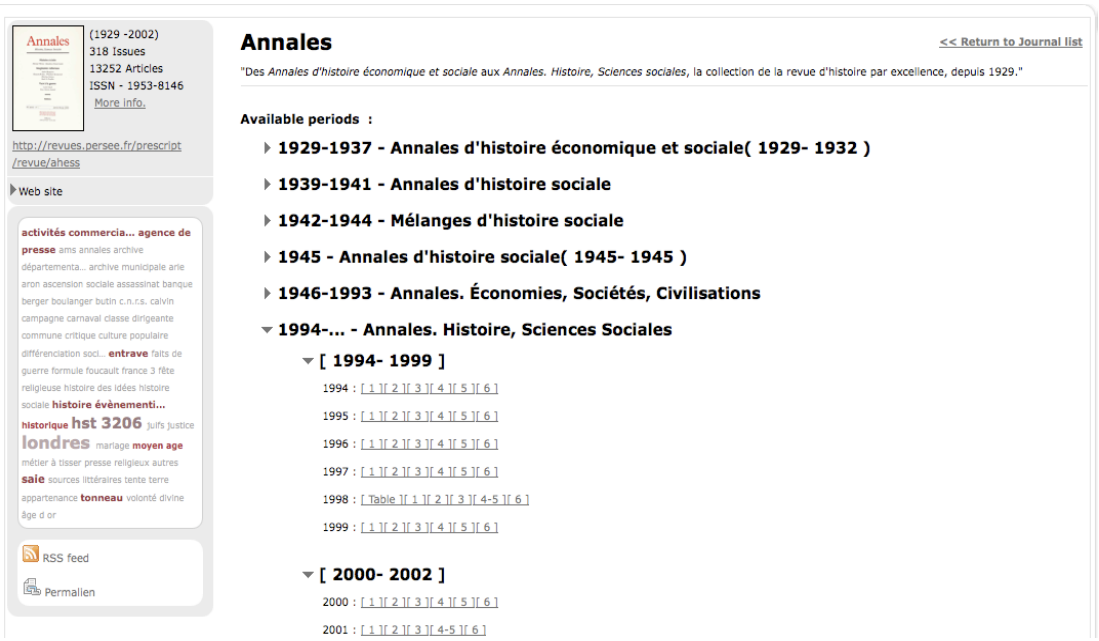


Figure 4 - Présentation de l'historique d'une revue

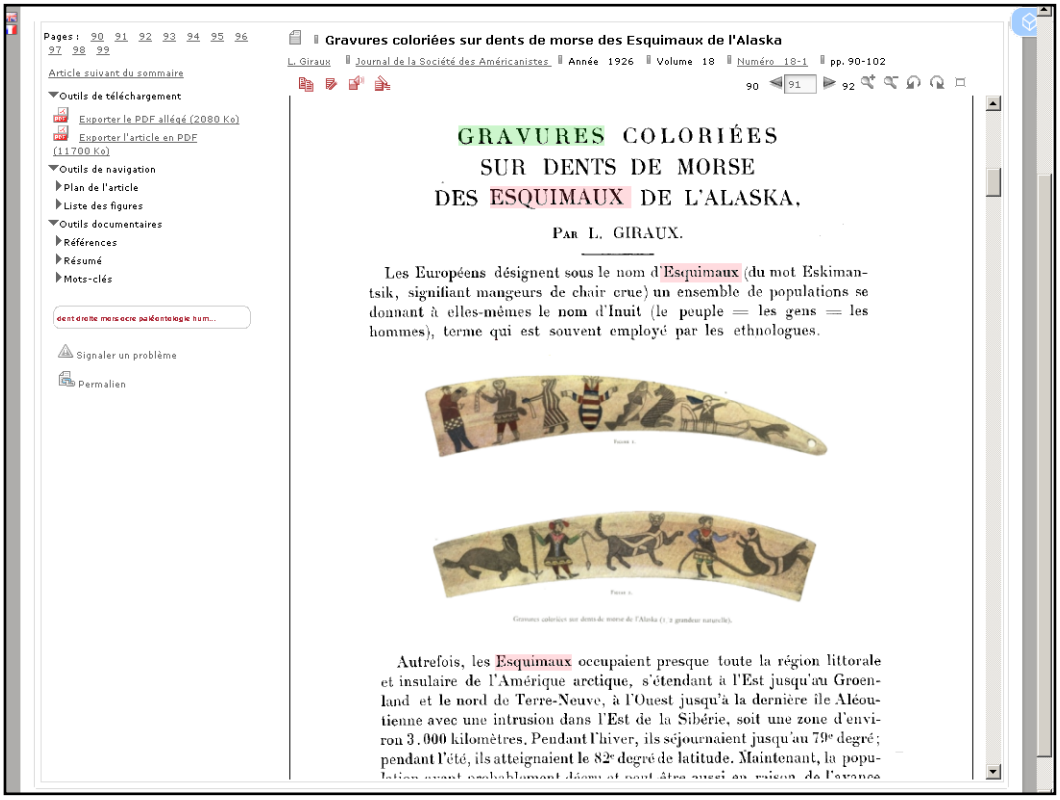


Figure 5 - Surlignement des termes recherchés dans l'image numérisée de l'article



Figure 6 - Plan et liste des illustrations d'un article

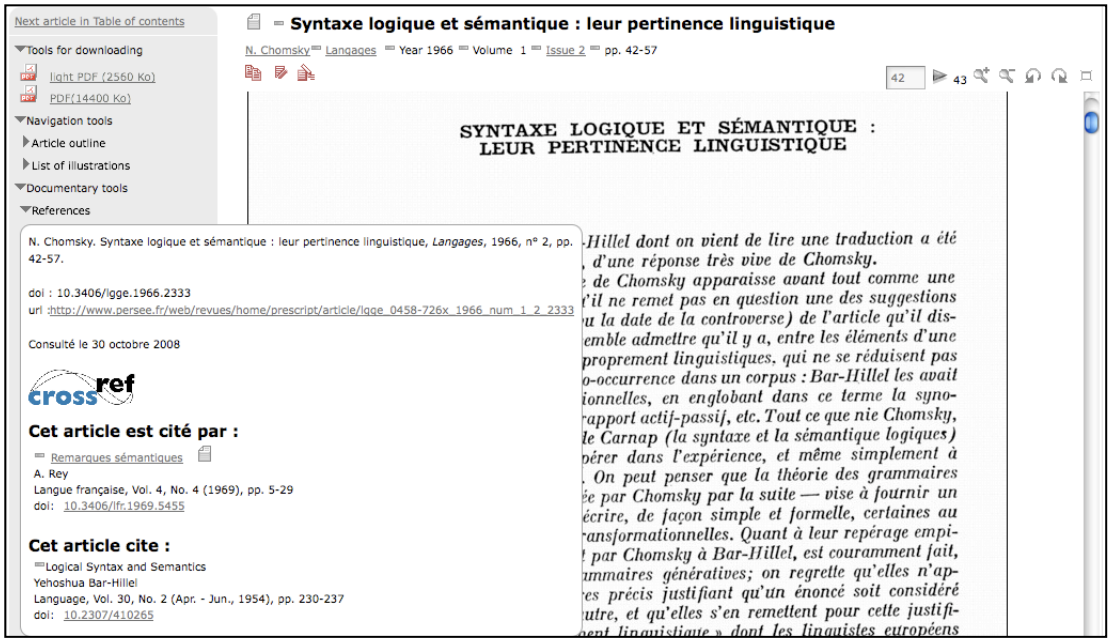


Figure 7 - Fonctionnalité de référencement croisé (Cet article est cité par / Cet article cite)

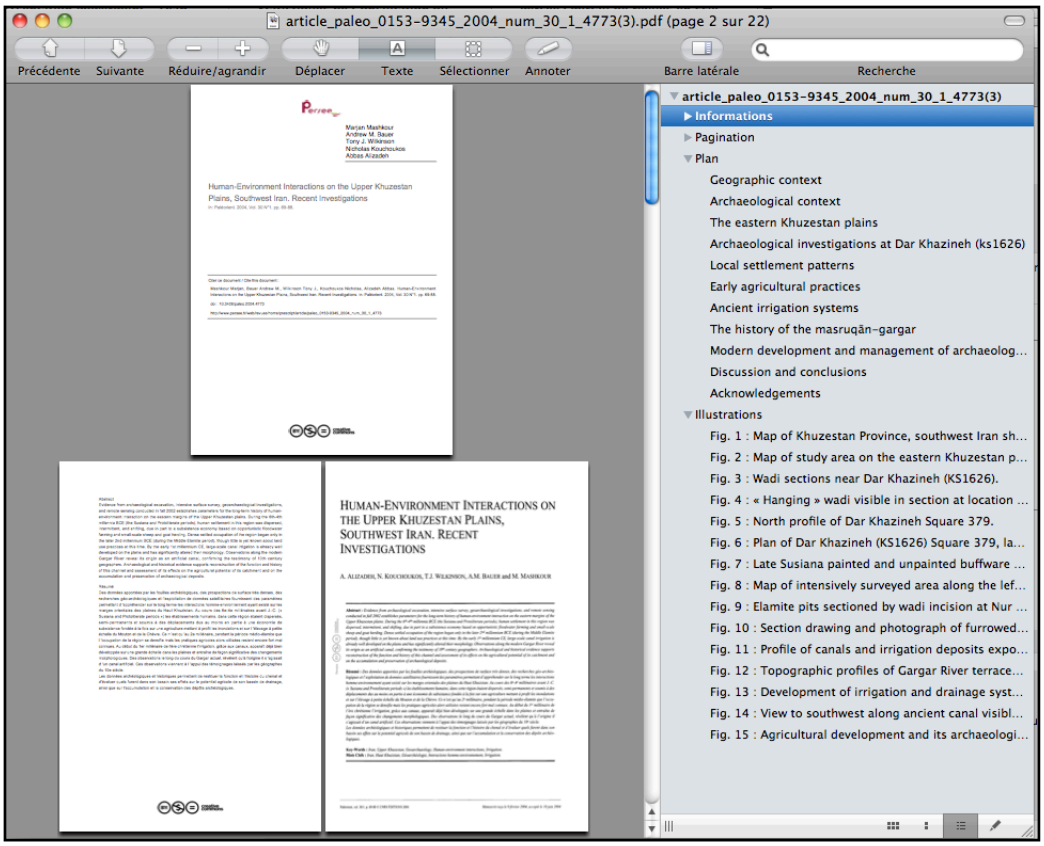


Figure 8 - Fichier PDF enrichi téléchargé à partir du port