



PERSÉE, partenaire confirmé

www.persee.fr

Infrastructure de recherche associée au dispositif CollEx, Persée développe et maintient une plateforme dédiée à la production et la diffusion de collections et de corpus documentaires numériques élaborés à partir de fonds patrimoniaux.

Au-delà de la mise à disposition de documents scientifiques, Persée vise à les intégrer pleinement dans un ensemble de connaissances (désormais mondialisé) en les rendant interopérables. Cela ne peut se faire qu'en mettant en œuvre des formats et des référentiels normalisés.

Persée n'a pas vocation à définir et maintenir de tels référentiels, aussi nous appuyons-nous sur des « produits » mis à disposition par des institutions partenaires. S'agissant des auteurs, parmi les référentiels envisageables, nous avons choisi celui proposé par l'Abes, IdRef, pour plusieurs raisons :

- son périmètre – issu du Sudoc, il couvre la plupart des auteurs scientifiques français

sur une très large période ;

- son ouverture – les données d'IdRef sont réutilisables, leur production et leur curation s'inscrivent dans un contexte international ;
- ses services – les différents outils – site web, API – facilitent l'exploitation et la mise à jour des données ;
- son dynamisme – IdRef bénéficie d'évolutions régulières, et son équipe est à l'écoute et réactive.

Depuis mars 2016, les liens entre les auteurs de Persée et IdRef sont établis soit au fil de la production – tout auteur « rencontré » au cours du processus de documentation est lié à une autorité IdRef et, si celle-ci n'existe pas, Persée peut la créer dans IdRef ; soit *a posteriori*, par un alignement automatique effectué grâce aux outils développés par l'Abes (Qualinca) ou par un documentaliste de l'équipe Persée ayant la possibilité d'alimenter IdRef (ajouts, compléments d'information, corrections, etc.).



En mars 2017, sur les 115 000 auteurs de Persée, la moitié sont liés à des autorités d'IdRef, ce qui permet dans un second temps de déduire de nouveaux liens vers d'autres référentiels (data.bnf.fr, VIAF, ISNI, dbPedia, Wikipedia, RePEc, etc.).

Outre la mise à disposition de données plus riches, d'outils de recherche, de parcours et de rebonds étoffés, les gestionnaires des différents systèmes d'information impliqués bénéficient d'une amélioration significative de la qualité des données primaires et des procédures automatisées de collecte d'informations complémentaires.

VIVIANE BOULÉTREAU

Responsable du pôle informatique et développements
UMS Persée

(Université de Lyon – ENS de Lyon – CNRS)
viviane.bouletreau@persee.fr

CONDITOR, le futur pot commun de métadonnées

www.bibliothequescientifiquenumerique.fr/conditor

Lancé officiellement le 5 décembre 2016, Conditor est un projet multipartenarial¹ porté par la DIST-CNRS et soutenu par le MENESR. Sa vocation : mettre en place un référentiel national de la production scientifique de la recherche publique française comportant :

- un « pot commun » de métadonnées alimenté par des sources multiples en provenance de l'enseignement supérieur et de la recherche (HAL, archives institutionnelles, bases bibliométriques, systèmes d'information recherche...), de sources internationales et de référentiels choisis collégialement ;
- un outillage adapté pour constituer ce réservoir et diffuser les métadonnées – validées ou à valider – vers les systèmes d'information de l'ESR.

En tant que référentiel national, Conditor se doit de garantir fiabilité et interopérabilité

sémantique entre ces sources bibliographiques diverses. Pour ce faire, il s'appuiera sur des référentiels communs adaptés et reconnus. Pour identifier les structures, le Répertoire national des structures de recherche (RNSR) sera utilisé. Pour identifier les auteurs, Conditor s'appuiera sur le référentiel IdRef porté par l'Abes. En effet, outre la richesse des données d'autorité produites par les réseaux de signalement (Sudoc, Thèses, Calames), les identifiants IdRef, grâce aux puissants algorithmes d'alignement développés au sein du projet de recherche Qualinca, sont alignés sur les identifiants IdHAL et Orcid, ce qui participe à l'objectif global d'enrichissement des corpus. Les apports prévus sont multiples : qualité et complétude des réservoirs institutionnels par le croisement de différentes sources, identification pérenne des acteurs (personnes

et structures), détection exhaustive de la production nationale, fiabilité des indicateurs de la science, facilitation du dépôt du texte intégral et des activités de gestion de la recherche pour les chercheurs...

D'ici deux ans, l'outillage nécessaire ainsi qu'un premier réservoir de métadonnées seront à disposition. Objectif à cinq ans : la mutualisation complète des travaux de référencement et l'intégration aboutie au sein des systèmes d'information.

ANNIE CORET

Responsable IST DIST/CNRS
annie.coret@cnrs-dir.fr

[1] Partenaires du projet : Abes, Amue, CCSD, Inist-CNRS, InSHS, Service d'appui à la politique et à la prospective scientifiques du CNRS, Huma-Num, Inra, Inria, Irstea, MENESR (équipes ScanR et RNSR), OST du HCERES, universités de Bordeaux, Paris-Dauphine, Paris-Diderot et UPMC.