

Un outil de comparaison visuelle des collections numérisées de Persée et de leurs équivalents imprimés



Deux étudiants du master Humanités numériques de Lyon ont comparé le corpus de revues numérisées par Persée et les revues imprimées présentes dans les Plans de conservation partagée des périodiques.

Le 31 janvier 2022 avaient lieu les soutenances de projets du master en humanités numériques, parcours *Pratiques et méthodes en Humanités numériques*, porté conjointement par l'université Lumière Lyon 2, l'Enssib, l'université Jean Moulin Lyon 3 et l'École normale supérieure de Lyon (cf p. 15). Parmi les quatre projets présentés, celui d'Amélie Descollonges et Julien Colin propose une visualisation graphique comparant les collections numérisées par Persée (463 revues numérisées) et leurs équivalents imprimés présents dans les établissements membres du réseau Sudoc. Précisons que, dans le cadre de ce projet, le jeu de données (liste de PPN)¹ a été fourni par l'Abes et que le code écrit par les étudiants, prochainement disponible via le GitHub de l'Abes, est réutilisable et rejouable à volonté.

UN TRAVAIL EN 2 VOLETS

En premier lieu, a été réalisée une rosace constituée de l'ensemble des titres Persée (1 barre = 1 titre), dans laquelle la hauteur de la barre représente le nombre de localisations de l'équivalent imprimé, et la couleur de la barre, la thématique disciplinaire de la revue.



Visualisation des localisations des périodiques numérisés par Persée répartis par thématiques

En second lieu, a été effectuée une comparaison titre à titre des états de collection de la revue papier et de la période numérisée par Persée.

Exemple avec le titre *Bulletin de l'Association française de recherche sur l'histoire du cinéma* : en vert les années 1986- 2000 auxquelles Persée donne accès ; en gris les états de collection de la revue imprimée dans les bibliothèques qui participent à un Plan de conservation partagée des périodiques. En glissant le curseur sur un des états de collection, on obtient plus d'informations sur le Plan concerné, les dates de début et de fin de l'état de collection et le nom de la bibliothèque de localisation.



Cette représentation sous forme de *timeline* se rapproche de celle de Périoscope², avec comme valeur ajoutée la présence de l'accès Persée, absente de Périoscope : pour la même revue, on aura accès aux états de collection et aux localisations des autres établissements Sudoc possédant le titre, mais n'émergeant à aucun Plan. Le nombre de localisations pouvant avoisiner les 400 pour certains titres du corpus, cette représentation textuelle complète avantageusement la représentation graphique précédente.

Un rapport d'erreur montrant les exemplaires papier sans état de collection est généré pour chaque titre.



UN COUP D'ACCÉLÉRATEUR POUR LA QUALITÉ DES DONNÉES

Ce travail a permis à Persée et à l'Abes de collaborer étroitement à l'amélioration de la qualité des données descriptives du corpus Persée. À partir du fichier KBart Persée disponible *via* la base de connaissance Bacon³, des vérifications d'identifiants ont été effectuées en interrogeant le Sudoc : correspondances entre les PPN et ISSN de Persée, les PPN et ISSN imprimés. Ce corpus d'environ 450 titres a également été comparé avec le corpus Persée signalé dans Mir@bel.

Ainsi, certaines erreurs ont pu être identifiées et corrigées : demande d'attribution d'ISSN papier et électroniques manquants, rectification du fichier KBart Persée (suppression des doublons, introduction des manquants, alignements d'identifiants).

LES PISTES DE TRAVAIL À VENIR ENTRE L'ABES ET PERSÉE

L'élan donné par la qualité du travail des étudiants devrait permettre de poursuivre l'amélioration réciproque de la qualité des données : établir davantage de passerelles entre les outils de l'Abes et Persée (qui utilise déjà IdRef) ; signaler dans le Sudoc les accès Persée ; améliorer la qualité des notices bibliographiques par des chantiers qualité sur le corpus de revues numérisées Persée (ajout d'indexation sujet, d'autorités collectivité ou auteurs, liens aux métarevues).

L'ÉQUIPE PROJET⁴

[1] *Pica production number* identifiant univoque pour chaque titre

[2] <https://periscope.sudoc.fr>

[3] <https://bacon.abes.fr>

[4] Julie Mistral et Morgane Parra, Service Ressources continues de l'Abes ; Sitthida Samath, responsable Qualité et Interopérabilité des Données de Persée ; avec l'aide d'Amélie Descollonges et Julien Colin, étudiants en master Humanités numériques de Lyon.