
Étude sur l'utilisation d'œuvres relevant des arts visuels dans les publications scientifiques

Plan de gestion de données créé à l'aide de DMP OPIDoR

Créateurs du PGD : Antoine Blanchard, Pierre-Carl Langlais

Affiliation du créateur principal : Other Organisation

Modèle du PGD : DCC Template (français)

Dernière modification du PGD : 25/02/2021

Financeur : Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Résumé du projet :

Alors que le droit d'auteur sur les images limite actuellement la diffusion de certains travaux issus de la recherche, le législateur a souhaité assurer la sécurité juridique des collectifs de recherche dans le respect des droits patrimoniaux et moraux attachés à ces œuvres : c'est l'objet de l'article 28 de la loi de programmation de la recherche. Cette étude vise à estimer le nombre d'images publiées entrant dans le champ de la mesure, et de fournir une méthodologie permettant de réitérer ce calcul sur les flux annuels de travaux et publications scientifiques.

Chercheur Principal : Pierre-Carl Langlais

Identifiant ORCID : 0000-0001-9035-1127

Contact pour les Données : Claire Leymonerie

Produits de recherche :

1. Base de métadonnées : Base de métadonnées des documents et images constituant le corpus de l'étude (Jeu de données)
2. Corpus de documents : Corpus de documents en accès ouvert servant de base à l'étude (Texte)
3. Corpus d'images : Corpus d'images en accès ouvert servant de base à l'étude (Image)
4. Document exécutable : Fichier .rmd exécutable sous R permettant d'extrapoler les résultats obtenus sur l'échantillon (Workflow)
5. Code source : Code source utilisé pour exécuter les différentes phases de l'étude (Logiciel)

Droits d'auteur

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque

relation avec votre projet ou votre soumission.

Étude sur l'utilisation d'œuvres relevant des arts visuels dans les publications scientifiques

Collecte des données

Base de métadonnées : Base de métadonnées des documents et images constituant le corpus de l'étude

En requêtant Isidore (outil proposé par Huma-Num), nous allons créer un corpus hybride comprenant :

- l'ensemble des publications de 2019 hors CAIRN (soit 277 500 références)
- les publications Cairn de 2017 (soit 17 000 références) plutôt que 2019 afin d'intégrer les publications ouvertes après expiration de la barrière mobile (en effet, les publications datées de 2019 sur Cairn sont majoritairement payantes et donc hors du champ de la mesure ; or il est important d'analyser également les pratiques de publication des revues disponibles sur Cairn, d'où le choix d'une année de publication plus ancienne car après 3 ans ces articles passent typiquement en accès ouvert)
- l'intégralité des publications disponibles dans Persée.

La présente base de métadonnées contiendra tous les champs permettant de décrire ce corpus de documents, et les images que nous allons en extraire. Les métadonnées seront fournies par Isidore ou renseignées directement par nous.

Les métadonnées seront obtenues par des requêtes sur l'API Isidore, complétées par des bases de données tierces : ISSN, Wikidata, RNSR...

Corpus de documents : Corpus de documents en accès ouvert servant de base à l'étude

En requêtant Isidore (outil proposé par Huma-Num), nous allons créer un corpus hybride comprenant :

- l'ensemble des publications de 2019 hors CAIRN (soit 277 500 références)
- les publications Cairn de 2017 (soit 17 000 références) plutôt que 2019 afin d'intégrer les publications ouvertes après expiration de la barrière mobile (en effet, les publications datées de 2019 sur Cairn sont majoritairement payantes et donc hors du champ de la mesure ; or il est important d'analyser également les pratiques de publication des revues disponibles sur Cairn, d'où le choix d'une année de publication plus ancienne car après 3 ans ces articles passent typiquement en accès ouvert)
- l'intégralité des publications disponibles dans Persée.

À partir de ce corpus, nous allons extraire (quand c'est possible) les documents disponibles en accès ouvert (sans paiement ni authentification).

Les données seront collectées depuis les plateformes de publication indexées par Isidore, par exemple OpenEdition, Cairn, HAL, TEL, Persée...

Corpus d'images : Corpus d'images en accès ouvert servant de base à l'étude

Une fois le corpus de documents constitué, nous allons extraire les images qu'ils contiennent en se limitant (dans la mesure du possible) aux illustrations utilisées à des fins scientifiques (en excluant notamment les logos par exemple en page de garde). Les images caviardées seront également présentes.

Dans le cas des documents HTML, les images seront extraites du code source (balise). Dans le cas des documents PDF, le package pdfimages sera utilisé.

Document exécutable : Fichier .rmd exécutable sous R permettant d'extrapoler les résultats obtenus sur l'échantillon

Nous allons produire un document exécutable, qui pourra être exécuté pour produire le document de synthèse en HTML ou en PDF contenant le résultat des scripts R, eux-mêmes inclus dans le code source du document exécutable. Cet outil fournira une estimation de la volumétrie globale et du nombre d'images entrant dans le champ de la mesure, tout en rendant chaque étape du calcul (échantillonnage, dénombrement, puis extrapolation) transparente et modifiable.

Le document sera créé en R markdown et exécutable dans le logiciel RStudio.

Code source : Code source utilisé pour exécuter les différentes phases de l'étude

L'ensemble des scripts utilisés au cours de cette étude sera consolidé dans ce jeu de données.

Les scripts seront créés dans RStudio.

Documentation et métadonnées

Base de métadonnées : Base de métadonnées des documents et images constituant le corpus de l'étude

Question sans réponse.

Corpus de documents : Corpus de documents en accès ouvert servant de base à l'étude

Question sans réponse.

Corpus d'images : Corpus d'images en accès ouvert servant de base à l'étude

Question sans réponse.

Document exécutable : Fichier .rmd exécutable sous R permettant d'extrapoler les résultats obtenus sur l'échantillon

Question sans réponse.

Code source : Code source utilisé pour exécuter les différentes phases de l'étude

Question sans réponse.

Aspects éthiques et juridiques

Base de métadonnées : Base de métadonnées des documents et images constituant le corpus de l'étude

Question sans réponse.

Ces données factuelles ne sont pas couvertes par le droit d'auteur.

Corpus de documents : Corpus de documents en accès ouvert servant de base à l'étude

Ces documents, publics et en accès ouvert, ne présentent pas d'enjeux éthiques.

Ces documents sont la propriété de leurs ayant droits, et leur diffusion par les auteurs de l'étude n'a pas été autorisée à ce stade.

Corpus d'images : Corpus d'images en accès ouvert servant de base à l'étude

Ces images, publiques et en accès ouvert, ne présentent pas d'enjeux éthiques.

Ces images, hors images du domaine public, sont la propriété de leurs ayants droits, et leur diffusion par les auteurs de l'étude n'a pas été autorisée à ce stade, hors images sous licence autorisant leur exploitation.

Document exécutable : Fichier .rmd exécutable sous R permettant d'extrapoler les résultats obtenus sur l'échantillon

Ces document exécutable ne présente pas d'enjeux éthiques.

Ce document exécutable, créé par les auteurs de l'étude, est placé sous licence Creative Commons CC-BY.

Code source : Code source utilisé pour exécuter les différentes phases de l'étude

Ce code source ne présente pas d'enjeux éthiques.

Ce code source, créé par les auteurs de l'étude, est placé sous licence Creative Commons CC-BY.

Stockage et sauvegarde

Les données sont stockées par Huma-Num sur l'espace mis à disposition de l'étude : <https://images-publications-scientifiques.huma-num.fr/>

Des droits d'accès ont été ouverts pour le seul Pierre-Carl Langlais.

Sélection et préservation

Au vu du potentiel de réutilisation des données pour reproduire l'étude ou pour informer d'autres travaux de recherche, elles devront être conservées pendant une durée de 10 ans sur l'infrastructure d'Huma-Num, sous la responsabilité de l'intendant des données du Ministère de l'enseignement supérieur, de la recherche et de l'innovation.

Aucune préservation à long terme n'est prévue.

Partage des données

Base de métadonnées : Base de métadonnées des documents et images constituant le corpus de l'étude

La base de métadonnées sera librement accessible sur le serveur Huma-Num, et sa découverte facilitée par un data paper.

Aucune restriction au partage des données n'est prévue.

Corpus de documents : Corpus de documents en accès ouvert servant de base à l'étude

Le corpus de document ne sera pas partagé.

Question sans réponse.

Corpus d'images : Corpus d'images en accès ouvert servant de base à l'étude

Le corpus de document ne sera pas partagé.

Question sans réponse.

Document exécutable : Fichier .rmd exécutable sous R permettant d'extrapoler les résultats obtenus sur l'échantillon

Le document exécutable sera librement accessible, et sa découverte facilitée par un data paper.

Aucune restriction au partage des données n'est prévue.

Code source : Code source utilisé pour exécuter les différentes phases de l'étude

Le code source sera librement accessible, et sa découverte facilitée par un data paper.

Aucune restriction au partage des données n'est prévue.

Responsabilités et ressources

Pierre-Carl Langlais et Antoine Blanchard sont responsables des données jusqu'au terme de l'étude. Cette responsabilité sera transférée au Ministère de l'enseignement supérieur, de la recherche et de l'innovation une fois l'étude terminée.

Question sans réponse.