



le **cnam**

COMITE DE SUIVI DE L'EDITION SCIENTIFIQUE

Janvier 2020

La diffusion sur Hal, Academia et ResearchGate des
articles de recherche des revues françaises de SHS

RAPPORT FINAL

Emile Gayoso, *Postdoctorant Dicen IdF (CNAM) - MESRI*

Table des matières

I.	Contexte de l'étude.....	4
II.	Méthodologie de la collecte de données	4
II.1	Etablir la liste des articles	5
II.2	Mesurer la présence des articles sur HAL.....	6
	De la revue à la discipline : constitution d'un tableau de bord de suivi des dépôts dans Hal	6
II.3	Mesurer la présence des articles sur Academia et ResearchGate.....	7
II.3.1	Échantillonnage.....	8
III.	Etude des dépôts dans Hal	8
III.1	Tableau synoptique des dépôts dans Hal	8
III.1.1	Notices et textes intégraux déposés sur la période 2010-2018	8
III.1.2	Notices et textes intégraux depuis la création de Hal.....	10
III.1.3	Analyse des taux de dépôt discipline par discipline	12
III.1.4	Focus sur les extrêmes : des revues aux cent dépôts aux revues sans- dépôts.....	14
III.1.5	Qui sont les déposants ?.....	16
III.2	Analyse dynamique des dépôts dans Hal.....	21
III.2.1	Evolution des dépôts sur la période 2010-2018.....	21
III.2.2	Définition et évolution du taux de dépôt des articles parus dans l'année	23
III.3	Délai de dépôt en archive ouverte des articles dans Hal et barrière mobile	24
III.3.1	Résultats au niveau de la revue	24
III.3.2	Résultats au niveau détaillé de l'article	25
	Barrière mobile et taux de dépôt	26
IV.	Etude de la diffusion des articles de recherche en SHS sur les réseaux sociaux académiques Academia et ResearchGate	27
	Bibliographie	29
	Annexes	30

Annexe 1 : Liste des revues, précédées de leur numéro dans le corpus général et de leur identifiant disciplinaire.....	30
Annexe 2 : Les modes de diffusion des revues françaises de SHS : quelques résultats d'une pré-enquête du Ministère de la Culture	35
Annexe 3 : Méthode d'extraction OAI-PMH des articles de revues diffusées sur OpenEdition avec R	37
Annexe 4 : Méthode d'extraction des données de dépôts sur HAL revue par revue.....	38
Annexe 5 : Tableaux de synthèse des dépôts dans Hal par discipline (2010-2018).....	40
Annexe 5.1 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Sciences de l'environnement ».....	40
Annexe 5.2 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Géographie ».....	41
Annexe 5.3 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Economie-gestion ».....	43
Annexe 5.4 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Histoire ».....	46
Annexe 5.5 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Littérature »	52
Annexe 6 : script Python pour l'identification de la présence des articles sur Academia et ResearchGate	57

I. Contexte de l'étude

En réponse à l'appel à manifestation d'intérêt du ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation visant à contribuer au suivi de la Loi pour une République numérique du 7 octobre 2016, l'équipe Dicen IdF a, dans le premier volet de sa réponse, proposé d'établir des indicateurs mesurant la disponibilité en libre accès des articles des revues de sciences humaines et sociales de droit français.

Ces revues sont au nombre de 368 ([L'annexe 1](#) en donne la liste exhaustive ainsi que la nomenclature établie pour l'étude) et ont été choisies par les membres du comité de suivi avec l'objectif de parvenir à des listes le plus exhaustives possibles pour les domaines disciplinaires suivants :

- L'économie-gestion : 66 revues.
- La géographie : 55 revues.
- L'histoire : 148 revues.
- La littérature : 86 revues.
- Les sciences de l'environnement : 13 revues.

Une grande majorité de ces revues dispose d'une diffusion en ligne de ses contenus (plus de 85% d'entre elles, réparties inégalement selon les disciplines, comme le précise le tableau 1 en [annexe 2](#)), et une large majorité d'entre elles passent pour ce faire par l'une des deux grandes plateformes de revues de sciences humaines et sociales françaises : CAIRN et OpenEdition¹. Plus précisément, voici comment se répartissent les revues selon le mode de diffusion en ligne choisi² :

	CAIRN	OpenEdition	Autres plateformes (Persée, JSTOR, etc.)	Site Web ad hoc	TOTAL LIENS DE DIFFUSION	TOTAL REVUES
Environnement	4	4	3	2	13	13
Géographie	26	29	5	2	62	55
Economie-gestion	44	10	12	5	71	66
Histoire	54	52	23	28	157	148
Littérature	16	28	29	19	92	86
TOTAL	144	121	71	56	392	368

Tableau 1 : Modes de diffusion en ligne des 368 revues du corpus réparties par discipline

II. Méthodologie de la collecte de données

Le premier travail, effectué par les membres du CSES, a consisté à établir la liste des revues pertinentes pour chaque discipline. Le principe qui a présidé à la constitution de ces listes a été celui

¹ Lorsqu'on considère la période de travail (2010-2018), presque toutes les revues sont diffusées en ligne soit sur CAIRN, soit sur OpenEdition, soit sur les deux. Les plateformes et sites comme Persée, JSTOR ou Gallica, jouent aussi un rôle de diffusion en ligne important pour les revues de l'échantillon du CSES, mais la plupart du temps, les années de parution qu'elles recouvrent sont antérieures à la période de travail. C'est pourquoi le nombre de revues mentionnées dans la colonne « Autres plateformes (Persée, JSTOR, etc.) » est si faible.

² On a pris en compte l'ensemble des liens de diffusion en ligne pour la période de travail 2010-2018, or une même revue peut être diffusée sur plusieurs plateformes, ce qui explique que le nombre total de liens (avant-dernière colonne) soit supérieur au nombre total de revues (dernière colonne).

de l'exhaustivité : les listes sont les plus étendues possibles et reflètent ainsi la diversité des modèles éditoriaux existant au sein de chacune de ces disciplines. On trouvera des revues disposant d'une diffusion de l'ordre de quelques dizaines d'abonnements, d'autres en ayant plusieurs centaines voire plusieurs milliers, des revues n'étant disponibles que sur leur site web et des revues utilisant plusieurs plateformes de diffusion, d'autres enfin n'ayant recours qu'au support papier ; des revues adossées à des laboratoires universitaires, d'autres à des associations professionnelles ou des sociétés savantes, d'autres encore à des maisons d'édition. La sélection des titres a été effectuée à partir de Journalbase (CNRS) et validée par un chercheur de la spécialité.

Après un débat en séance avec les membres du CSES, la période choisie pour le travail d'établissement des indicateurs de présence en ligne des articles de ces revues a été **2010-2017**, laquelle a **ensuite été étendue à 2010-2018** lorsqu'il a été possible, courant 2019, d'actualiser les données avec les articles parus en 2018. **Ce choix permettait d'avoir suffisamment de profondeur temporelle pour remonter au-delà des temps d'embargo les plus longs parmi ceux ordinairement pratiqués (de 1 à 5 ans) et de disposer de deux années complètes (2017 et 2018) après la promulgation de la Loi pour la République numérique dont il s'agit ici de mesurer l'impact.**

II.1 Etablir la liste des articles

Les contenus des revues académiques sont très divers et les prendre tous en considération expose à faire porter l'analyse sur des objets peu comparables. L'architecture en rubriques des revues académiques peut en effet être très complexe et varier sensiblement d'une revue à l'autre. Afin de limiter les effets de cette variation, nous avons exclu les éditoriaux, les articles d'actualité, les introductions de dossier, les notes de lecture, les entretiens et de façon générale toute rubrique idiosyncrasique ne contenant pas d'articles de recherche³.

Puisque la période de référence choisie est 2010-2018, nous avons établi la liste des articles parus entre 2010 et 2018 dans les revues des cinq domaines disciplinaires retenus. Pour ce faire, nous avons procédé de trois manières différentes, selon les outils et services qui étaient à notre disposition respectivement pour les revues principalement diffusées par CAIRN, OpenEdition ou un autre site ou plateforme. **Pour les revues du portail CAIRN**, nous avons fait une demande auprès de la société pour obtenir directement la liste des publications sur la période concernée. **Nous avons alors obtenu un tableau de données au format .xlsx contenant 128 101 lignes correspondant chacune à une publication d'une des revues de CAIRN.** Toutes les publications se trouvaient dans ce fichier et nous avons dû procéder à un filtrage semi-automatisé des articles pour ne conserver que ceux correspondant à la définition donnée plus haut⁴.

Pour les revues principalement diffusées par OpenEdition, nous avons procédé autrement. Nous avons été invités par la société à extraire les données grâce à l'entrepôt OAI-PMH **rendant accessible l'intégralité des métadonnées aux formats METS et Dublin Core.** Nous avons alors utilisé un

³ Précisons que la mesure de la présence en ligne de la recherche française en sciences humaines et sociales dans sa globalité aurait imposé de prendre en compte toutes les revues, françaises et étrangères, dans lesquelles les chercheurs publient, et idéalement l'ensemble de leurs productions : communications, posters, rapports, *working papers*, etc. Une telle entreprise est au-delà de la mission du CSES qui enquête sur les revues de SHS et partant sur ce format spécifique (et principal !) de la communication scientifique qu'est l'article.

⁴ Pratiquement, nous avons éliminé automatiquement de la base de données de CAIRN les articles inférieurs à 5 pages et nous avons ensuite opéré un tri à la main de ceux restants afin de vérifier qu'il s'agissait bien d'articles de recherche comme nous les avons définis.

protocole d'extraction des métadonnées avec le logiciel R établi par un chercheur autrichien de l'*Institute of Statistics and Mathematics* de Vienne (Hornik, 2017). Celui-ci (qu'on trouvera reporté en annexe 3) permet de **lister l'intégralité des publications des revues choisies** et il a fallu à nouveau mettre en place un protocole semi-automatisé de nettoyage des données pour **ne conserver que les « articles » au sens de notre étude.**

Enfin, **pour les revues diffusées sur Persée, JSTOR**, sur un site propre ou seulement en version papier, nous avons **établi à la main les tableaux de titres d'articles correspondant à la période 2010-2018 en choisissant dès le report des données de ne conserver que les titres des articles de recherche, leur date de publication et leurs auteurs.**

II.2 Mesurer la présence des articles sur HAL

Créée en 2000 au sein de l'Institut national de physique nucléaire et de physique des particules (IN2P3), la plateforme de dépôt de documents scientifiques Hal est développée dès 2001 par le Centre de Communication scientifique Directe (CCSD) du CNRS et devient en 2006, suite à un protocole d'accord signé en 2006 par les universités et grandes écoles françaises à l'académie des sciences, la plateforme commune d'archivage de la recherche scientifique des établissements d'enseignement supérieur français.

En décembre 2018, elle compte **plus de 1 650 000 références de documents scientifiques⁵, parmi lesquels un tiers sont archivés en texte intégral (environ 550 000).** L'essentiel des dépôts est ainsi fait de notices contenant les métadonnées de documents scientifiques : titre du document, nom de l'auteur, date de parution, nom de la revue, etc. Hal est ainsi utilisé à la fois comme un outil de « communication scientifique directe » rendant accessible des articles, publiés ou non dans des revues, et un outil de bibliométrie permettant aux universités et institutions de recherche de mesurer la production scientifique de leurs laboratoires.

Quoi qu'il en soit, HAL, en tant qu'archive ouverte nationale en France, est un dispositif incontournable pour qui cherche à mesurer l'essor de l'open access dans les sciences sociales françaises. Nous avons donc mesuré, discipline par discipline et revue par revue, tout au long des 9 années concernées par la période (2010-2018) la nature et l'évolution des dépôts effectués sur la plateforme par les chercheurs eux-mêmes ou par un tiers. Le détail de la méthode d'extraction des données à partir de Hal se trouve en [Annexe 3](#).

De la revue à la discipline : constitution d'un tableau de bord de suivi des dépôts dans Hal

Une fois extraites les métadonnées⁶ des dépôts de chacune des 365 revues, on constitue une base de données année par année et pour chaque revue du nombre de dépôts réalisés. A titre d'exemple,

⁵ Parmi ces documents scientifiques, on compte une majorité d'articles (32 % des dépôts), suivis des actes de colloques (22 %), des chapitres d'ouvrage (19 %) ; viennent ensuite les photos, les *working papers*, les thèses, etc.

⁶ Parmi ces métadonnées figurent le nom de la revue, son éditeur, le titre de l'article, le nom de l'auteur, la date de parution, mais aussi la date de dépôt de l'article (qui peut être aussi bien la date de dépôt d'une notice transformée par la suite en texte intégral que la date de dépôt d'un texte intégral), le nom complet du déposant (qui n'est pas nécessairement l'auteur ni même une personne physique), la date de libération du

pour les revues du domaine disciplinaire « sciences de l'environnement », voici le tableau correspondant :

Nom de la revue	Nombre de dépôts d'articles (parus entre 2010 et 2018) en texte intégral pour les revues de sciences de l'environnement									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	
Développement durable et territoires	0	0	6	2	7	5	6	19	3	
ELOHI	0	0	0	0	0	0	3	0	0	
Environnement, risques et santé	0	0	0	0	0	0	0	0	0	
Natures Sciences Sociétés : recherches et débats interdisciplinaires	0	6	2	2	6	12	7	8	10	
Revue d'écologie	0	2	0	0	1	0	1	3	3	
Revue juridique de l'environnement	0	0	1	0	1	0	3	0	7	
SAPIENS (Surveys and Perspectives Integrating Environment and Society)	0	0	0	0	0	0	0	0	0	
Water Alternatives	0	0	0	0	1	0	0	1	0	
Revue d'ethnoécologie	0	0	0	0	1	3	3	2	1	
Annales des Mines - Responsabilité et environnement	1	0	1	0	1	0	0	1	2	
Ecologie & politique	0	0	0	0	0	1	0	1	1	
Tous urbains	0	0	0	0	0	0	1	0	1	
Climatologie	0	1	0	1	0	1	3	0	1	

Tableau 2 : Exemple de tableau de bord de suivi des dépôts dans Hal pour les revues de la discipline "Environnement"

Ces données, rapportées à celles sur le nombre d'articles de recherche publiés chaque année par les revues, que nous avons également établies (cf. la section ci-avant : [Etablir la liste des articles](#)), permettent de mesurer l'importance relative prise par les dépôts, la nature de ces derniers (notices ou textes intégraux) et de suivre ces indicateurs de façon dynamique de 2010 à 2018.

II.3 Mesurer la présence des articles sur Academia et ResearchGate

Academia et ResearchGate sont des sites de réseaux sociaux créés tous deux en 2008 aux Etats-Unis respectivement à San Francisco et à Boston. Academia a été créé par Richard Price, docteur en philosophie de l'Université d'Oxford et entrepreneur, tandis que ResearchGate est le fruit de la collaboration entre deux virologues (Ijad Madisch et Sören Hofmayer) et un informaticien (Horst Fickenscher). Ces sites de réseaux sociaux permettent aux chercheurs de se créer des profils académiques dotés d'informations sur leur parcours, voire sur leur vie personnelle, mais surtout ils autorisent la diffusion des textes intégraux des publications numériques des chercheurs (articles, chapitres d'ouvrage, posters, présentations powerpoint, etc.). Grâce à leurs robots d'indexation parcourant le web et en particulier les bases de données des plateformes des éditeurs (Cairn, OpenEdition, JSTOR, pour nous), ces entreprises fonctionnant comme des *startup* (leur modèle économique est pour l'essentiel basé sur les levées de fonds) peuvent répertorier les publications des auteurs ayant créé un compte sur leur site et leur suggérer de diffuser les publications ainsi inventoriées. Elles jouent ainsi un rôle d'incitation à la diffusion des textes intégraux de leurs productions par les chercheurs, ce qui rend leur étude indispensable au moment de mesurer la diffusion des articles de recherche sur Internet en dehors des sites et plateformes des éditeurs.

dépôt, la présence ou non d'un texte intégral (ou seulement d'une notice), la présence ou non d'un lien hypertexte vers un texte intégral en accès libre, la langue du document, etc.

Le site de réseau social académique Academia.edu ne rendant pas disponible les métadonnées de ses dépôts, nous avons opté pour une stratégie de *web scrapping* consistant à programmer un script permettant d'automatiser des requêtes dans la base d'articles d'Academia. Ainsi, à partir des tableaux contenant les titres des articles publiés de 2010 à 2018, **le script est capable d'extraire chaque titre d'article et de déclencher une requête dans la base de données d'Academia pour vérifier sa présence éventuelle**. Si l'article est présent, le script renvoie un 1 dans une colonne supplémentaire qu'il crée au sein du tableau de sortie ; si l'article est absent, il renvoie un 0.

II.3.1 Échantillonnage

Etant donné le volume d'articles parus de 2010 à 2018 dans la totalité du corpus de revues, l'obtention de résultats revue par revue et article par article est longue et fastidieuse car elle demande de lancer le script pour chacun des articles paru dans chacune des revues. Une méthode alternative consiste à procéder par échantillonnage. C'est ce que nous avons fait pour chaque discipline. Nous avons ainsi réalisé, à partir des bases de données d'articles parus de 2010 à 2018, des échantillons aléatoires représentatifs de la totalité des articles concernés. En choisissant une marge d'erreur de 5% et un intervalle de confiance de 95%, on obtient $n = 384,16$, **soit une taille d'échantillon arrondi à l'entier supérieur de 385 articles**⁷.

III. Etude des dépôts dans Hal

III.1 Tableau synoptique des dépôts dans Hal

III.1.1 Notices et textes intégraux déposés sur la période 2010-2018

On s'intéresse d'abord aux dépôts effectués sur la période de travail : 2010-2018. Lorsqu'on somme tous les dépôts effectués entre 2010 et 2018 pour des articles parus entre 2010 et 2018, et que l'on décline les totaux disciplines par disciplines, on obtient le tableau suivant :

	Nombre d'articles en texte intégral déposés	Nombre de notices déposées	Total des dépôts par discipline	Pourcentage de dépôts d'articles en texte intégral par rapport au total de dépôts (en %)
Environnement	178	374	552	32,2

⁷ On utilise la formule $n = z \times (p(1-p)) / m^2$, avec n qui est la taille de l'échantillon, z le niveau de confiance selon la loi normale centrée réduite, p la proportion estimée de la population présentant la caractéristique étudiée (ici la caractéristique étudiée est la présence de l'article sur un site de réseau social académique). Comme on ne connaît pas cette fréquence, qui est l'objet même de l'étude, on choisit la valeur de « p » qui maximise le numérateur, maximisant ainsi la taille de l'échantillon. On peut montrer que c'est $p=0,5$ qui remplit ces conditions. m : marge d'erreur jugée acceptable pour la détermination de la fréquence de l'attribut (ici, on choisira 5%) Avec un intervalle de confiance à 95% et un $p=0,5$ et $m=0,05$, on obtient $n=384,16$, soit une taille d'échantillon arrondi à l'entier supérieur de 385 articles.

Eco-gestion	1117	4031	5148	21,7
Géographie	929	2733	3662	25,4
Histoire	759	2844	3603	21,1
Littérature	390	1577	1967	19,8
Total des dépôts par type de document	3373	11559	14 932	22,6

Tableau 3 : Nombre de documents et notices déposés sur Hal pour les 368 revues du corpus sur la période 2010-2018

On remarque immédiatement que **les dépôts de textes intégraux sont bien moins nombreux que les dépôts de notices**. Plus précisément, ils représentent 3373 documents sur les 14 705 répertoriés au total, soit **un peu moins de 23% des dépôts**. On constate par ailleurs quelques différences disciplinaires, puisque pour la géographie et les sciences de l'environnement cette proportion est plus proche d'1 dépôt sur 3 (30%) tandis qu'en littérature, histoire et économie-gestion, elle équivaut à 1 dépôt sur 5 (20%).

Mais **ces chiffres ne prennent tout leur sens qu'une fois comparés à ceux des parutions d'articles dans les revues** qui constituent en quelque sorte la source du flux de dépôts que nous mesurons sur Hal. Pour ce faire, nous avons procédé à un dénombrement manuel des articles de recherche parus dans les revues du corpus, lequel nous a permis d'estimer pour chaque discipline le nombre total d'articles publiés sur la période 2010-2018 en complétant les données manquantes pour certaines revues en 2018⁸ par des projections s'appuyant sur les chiffres établis pour les années précédentes.

	Nombre de revues pour chaque discipline	Ordre de grandeur du nombre d'articles parus de 2010 à 2018	Estimation de l'intensité annuelle de parution pour chaque revue (en nombre d'articles de recherche par an)
Environnement	13	3500	30
Eco-gestion	66	15 000	25
Géographie	55	11 500	23
Histoire	148	25 000	19
Littérature	86	15 000	19,5
TOTAL	368	70 000	

Tableau 4 : Nombre d'articles parus dans les 368 revues du corpus sur la période 2010-2018

Un premier constat s'impose : l'ordre de grandeur du nombre d'articles parus dans la totalité des revues du corpus et celui des dépôts effectués n'est pas le même. On avait en effet un peu moins de 15 000 dépôts dans Hal dont moins de 3 500 textes intégraux, alors qu'on a environ **70 000 articles sur la période 2010-2018 dans les revues du corpus**.

On mesure par ailleurs les différences disciplinaires en matière d'intensité de parution dans les revues scientifiques. Les sciences de l'environnement et l'économie-gestion sont les disciplines dans lesquelles les revues publient en moyenne le plus d'articles de recherche (respectivement 30 et 25 par an) tandis que les revues d'histoire et de littérature en publient environ 19 et celles de géographie 23.

⁸ En effet, le dénombrement a été effectué à l'été 2019, date à laquelle il restait une trentaine de revues environ à n'avoir pas publié leur dernier numéro de 2018.

On peut maintenant confronter les données de dépôts et celles de parution pour établir les taux de dépôts d'articles en texte intégral par discipline :

	Ordre de grandeur du nombre d'articles parus de 2010 à 2018	Nombre d'articles en texte intégral déposés	Taux de dépôt d'articles en texte intégral sur la période 2010-2018 (en %)	Nombre moyen de dépôts d'articles par an pour une revue
Environnement	3500	178	5,8	≈ 1,5
Eco-gestion	15 000	929	6,8	≈ 1,6
Géographie	11 500	1117	10,0	≈ 2,2
Histoire	25 000	759	2,7	≈ 0,5
Littérature	15 000	390	2,6	≈ 0,5

Tableau 5 : Estimation du nombre moyen de dépôts de textes intégraux déposés chaque année par revue au sein de chaque discipline

Les taux de dépôts d'articles varient d'un peu moins de 3% pour la littérature et l'histoire à près de 10% pour la géographie en passant par environ 5% et 6% respectivement pour l'environnement et pour l'économie-gestion. Pour rendre ces taux plus parlants, on peut les rapporter à ce qu'ils représentent en termes de nombre de dépôts d'articles moyen par an et par revue. On voit que les résultats sont très faibles puisque **les revues d'histoire et de littérature se caractérisent par en moyenne un dépôt de texte intégral tous les deux ans** (0,5 par an), celles **d'économie et de sciences de l'environnement par environ 3 dépôts tous les deux ans** (1,5 par an), quand **la géographie arrive en tête avec ses 2,2 articles déposés par revue et par an**.

Comparons maintenant ces premiers résultats sur la période 2010-2018 à ce que l'on obtient, pour le même corpus, sur l'intégralité des dépôts disponibles depuis le lancement de Hal en 2001.

III.1.2 Notices et textes intégraux depuis la création de Hal

Si Hal est lancé en 2001, les premiers dépôts sont répertoriés en 2002. L'étude de l'ensemble des dépôts réalisés pour les revues du corpus recouvrent donc 17 années, entre 2002 et 2018 incluses⁹.

	Nombre de documents déposés	Nombre de notices déposées	Nombre de notices déposées avec lien vers texte intégral en OA	Pourcentage de dépôts d'articles en texte intégral sur la totalité des dépôts (en %)
Environnement	254	642	110	28,3
Eco-gestion	2014	7185	277	21,9
Géographie	2245	4897	556	31,4
Histoire	1700	4888	316	24,8
Littérature	676	2641	347	20,4
TOTAL	6889	20223	1606	25,2

Tableau 6 : Nombre de documents et notices déposés sur Hal pour les 368 revues du corpus sur la période 2002-2018

⁹ Si l'on ne dispose des titres d'articles parus que sur la période 2010-2018, on a en revanche extrait les données de dépôt sur Hal correspondant aux 368 revues du corpus sur toute la période disponible, c'est-à-dire depuis la mise en service de Hal en 2001 et les premiers dépôts constatés, ici en 2002.

Ce tableau permet d'avoir un premier aperçu de la différence entre la période récente (2010-2018) et la période complète (2002-2018) en montrant que globalement **la proportion de textes intégraux parmi les dépôts sur Hal n'a pas augmenté sur la première**. En effet, hormis l'exception des sciences de l'environnement, les quatre autres disciplines ont des profils identiques où les notices prennent plus d'importance sur la période 2010-2018 qu'elles n'en ont sur la période 2002-2018. Cela signifie que **les chercheurs des SHS françaises et les personnels des services tiers en charge des dépôts pris ensemble ont eu davantage tendance, au cours de la dernière décennie, à déposer les métadonnées des articles que les textes intégraux de ces derniers**.

Regardons désormais comment évoluent les dépôts sur toute la période :

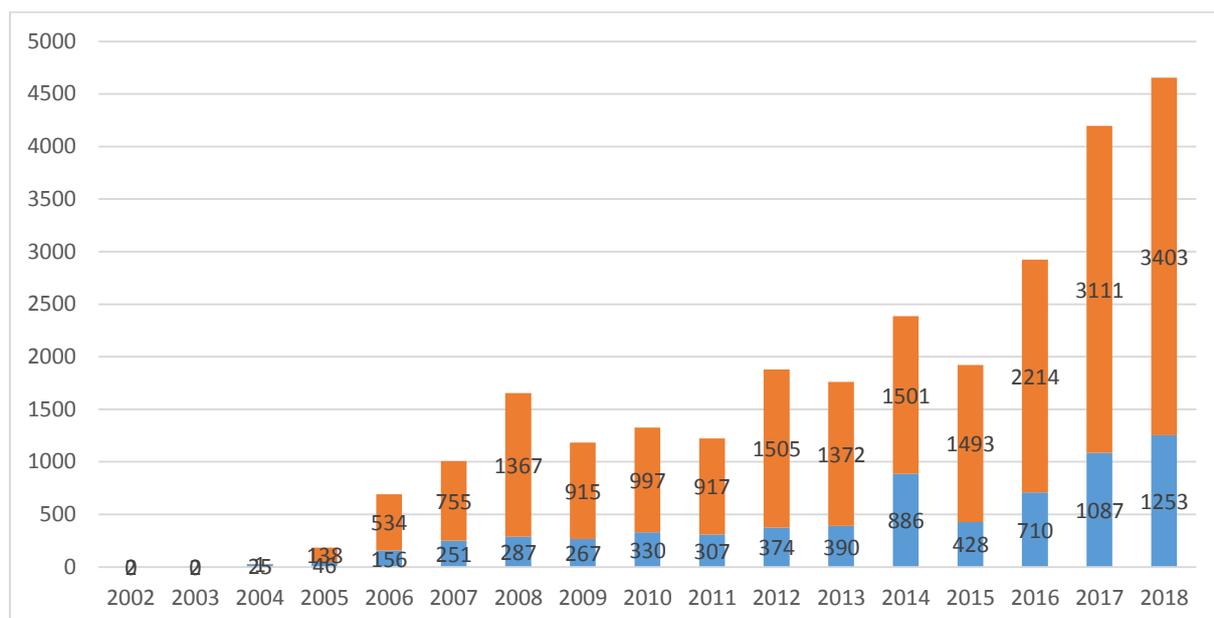


Figure 1 : Nombre de dépôts de textes intégraux et de notices sur le corpus de 368 revues (2002-2018)

Après les balbutiements de 2002 et 2003, une progression linéaire de quelques dizaines de dépôts par an s'amorce entre 2004 et 2013. Un changement d'échelle s'opère en 2014 où le nombre de dépôts est plus que doublé par rapport à 2013. L'année 2015 semble contredire ce changement d'échelle, mais il se confirme pourtant avec l'année 2016 et la rupture de pente est entérinée dès lors avec une progression annuelle non plus de l'ordre de la dizaine de dépôts de textes intégraux mais de l'ordre de la centaine.

Concernant les dépôts de notices, ceux-ci commencent seulement en 2004 avec 1 notice déposée pour tout le corpus et décollent en 2005 avec 138 notices déposées cette année-là. Les dépôts suivent une progression linéaire d'environ 200 dépôts annuels jusqu'en 2008, année où les dépôts doublent par rapport à 2007, passant de 755 à 1367. Les années 2009, 2010, 2011 voient une régression du nombre de notices déposées qui ne dépasse pas 1000. En 2012 néanmoins, on atteint les 1505 notices déposées et ce niveau va se maintenir en un palier jusqu'en 2015 inclus. **A partir de 2016, le dépôt de notices progresse très fortement passant de 2214 cette année-là à 3403 en 2018 en passant par 3111 en 2017.**

Intéressons-nous désormais au ratio dépôts de textes intégraux/dépôts de notices. Celui-ci évolue comme suit :

Année	2002	2003	2004	2005	2006	2007	2008	2009	2010
-------	------	------	------	------	------	------	------	------	------

Ratio textes intégraux/notices	ND ¹⁰	ND	25	0,333	0,292	0,332	0,21	0,292	0,331
Année	2011	2012	2013	2014	2015	2016	2017	2018	
Ratio textes intégraux/notices	0,335	0,249	0,284	0,59	0,287	0,321	0,349	0,368	

Les premières années, jusqu'en 2004, ne sont pas significatives car le dépôt de notices y est quasiment nul. Ce qu'il est intéressant de constater, c'est l'étonnante stabilité de ce ratio qui, entre 2005 et 2016, tourne le plus souvent autour d'1/3. Deux exceptions sont notables néanmoins sur cette période : 2008 et 2014, l'une avec un ratio de 1 texte intégral pour 5 notices environ (0,21) et l'autre avec près de 2 textes intégraux pour 3 notices (0,59). La tendance qui se dégage dans les quatre dernières années est celle d'une progression linéaire mais lente de ce ratio, passé de 0,287 à 0,368 entre 2015 et 2018.

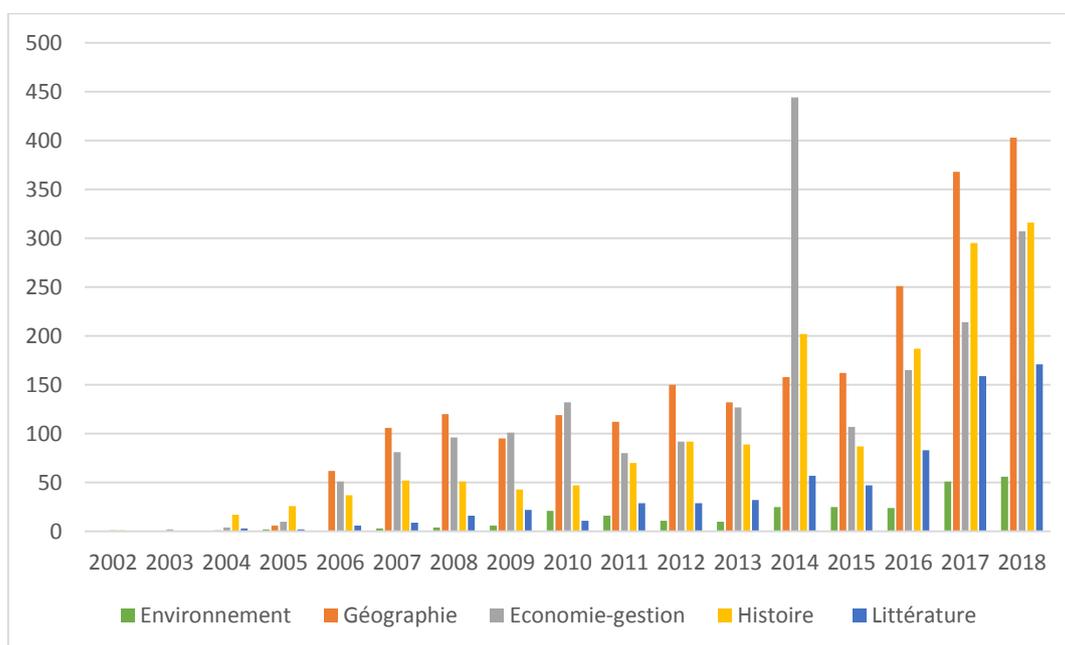


Figure 2 : Evolution du nombre de dépôts d'articles par disciplines sur la période 2002-2018

III.1.3 Analyse des taux de dépôt discipline par discipline

Le premier constat qui doit être fait est celui de la faiblesse des dépôts en libre accès sur Hal. En effet, le taux moyen de dépôt pour une discipline¹¹ le plus élevé est celui de la géographie et il plafonne à 9,8% (cf. diagramme box plot ci-dessous) tandis que les taux moyens de dépôts obtenus pour les sciences de l'environnement et l'économie-gestion sont respectivement de 5,7% et 6,4% ;

¹⁰ ND : Non Défini car aucune notice n'a été déposée.

¹¹ En rapportant le nombre d'articles déposés sur la période 2010-2018 au nombre d'articles parus dans une revue, on peut calculer un taux de dépôt par revue. En faisant la moyenne de ces taux pour toutes les revues d'une discipline, on obtient un taux moyen de dépôt dans Hal par discipline sur la période 2010-2018.

viennent enfin la littérature et l'histoire avec des taux de dépôt moyen respectivement de 3,1% et de 2,7%¹².

Ce constat général masque bien entendu de grandes disparités dont le diagramme *box plot* permet de mesurer l'ampleur. Ce type de diagrammes permet en effet de représenter la distribution des valeurs prises par les taux de dépôts pour chaque revue au sein des trois ensembles disciplinaires considérés.

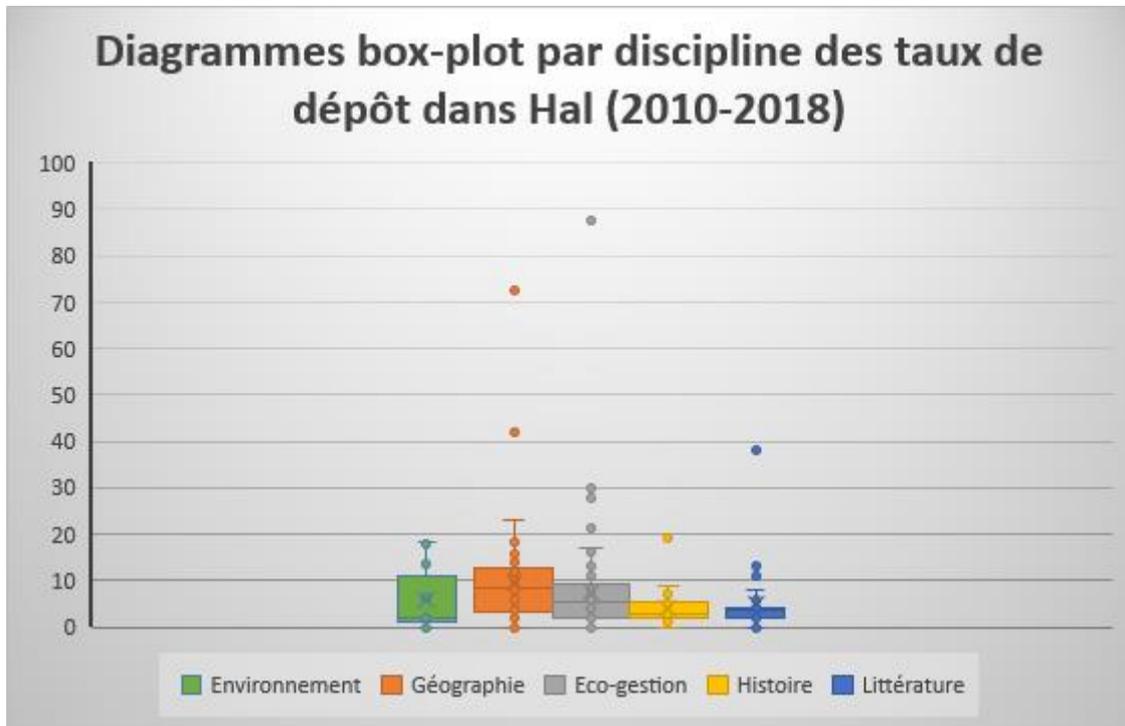


Figure 3 : Diagrammes box-plot par discipline des taux de dépôt dans Hal (2010-2018)

Pour la géographie et l'économie-gestion, les valeurs moyennes du taux de dépôt sont fortement tirées par quelques valeurs extrêmes très élevées. En particulier, deux revues se distinguent par des taux de dépôt supérieurs à 50% : la *Revue d'Étude en Agriculture et Environnement (Review of Agricultural, Food and Environmental Studies)* en économie-gestion (87,5% de dépôts dans Hal) et la revue *Recherche Transports Sécurité* en géographie (72,5% de dépôts dans Hal). Or, ces deux revues ont des politiques fortes en matière de dépôt dans les archives ouvertes puisqu'elles ont choisi d'en faire un moyen de diffusion premier de leurs productions, ce qui est suffisamment rare pour être souligné. La première dispose d'un portail bilingue sur Hal et d'un autre sur RePEc¹³ mettant l'intégralité de ses numéros en ligne jusqu'en 2015 (les trois dernières années étant en accès payant sur la plateforme Springer) tandis que la seconde, adossée à l'Ifsttar (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux), a opté pour la diffusion en open access immédiat depuis Hal à partir de 2018¹⁴.

¹² Pour ces deux dernières disciplines, les pourcentages fournis seront à consolider avec les données complètes car ils ont été calculés à partir d'un échantillon d'une trentaine de revues de chacune de ces deux disciplines.

¹³ RePEc, pour *Research Papers in Economics*, est une plateforme d'archives ouvertes internationale créée en 1997 à la suite du projet NetEc ayant lui commencé en 1993 et spécialisée dans les sciences économiques.

¹⁴ On peut ainsi lire sur la page web de la revue sur site de l'Ifsttar : « En cohérence avec son engagement pour l'Open Science, l'Ifsttar a choisi, à partir de 2018, de diffuser la revue RTS exclusivement en édition numérique et en open access sur l'archive ouverte HAL. »

III.1.4 Focus sur les extrêmes : des revues aux cent dépôts aux revues sans-dépôts

Les revues présentant un grand nombre de dépôts sont rares dans le corpus. Si l'on place la barre à 100 dépôts sur la période de travail, on n'en compte que 2 : Recherche Transport Sécurité (G_045) et la Revue de l'OFCE (EG_048). En descendant cette barre à 50 dépôts, on obtient une liste de 7 revues très inégalement répartie sur le corpus :

Environnement	Développement durable et territoires	53
Environnement	Natures Sciences Sociétés : recherches et débats interdisciplinaires	55
Géographie	Cybergeo	71
Géographie	Population & Avenir	68
Géographie	Recherche Transports Sécurité	147
Eco-gestion	Revue de L'OFCE	137
Eco-gestion	Revue Française de Gestion	51

Tableau 7 : Liste des revues présentant plus de 50 dépôts sur la période 2010-2018

Par symétrie, l'histoire et la littérature sont fortement pourvoyeuses de revues sans dépôts, au point qu'il convient d'en faire une catégorie à part entière des revues de ces deux dernières disciplines, ce qui ne se justifie pas pour les trois autres :

- Environnement : 1
- Géographie : 1
- Economie-gestion : 1
- Histoire : 26
- Littérature : 18

On trouvera dans le tableau ci-joint la **liste des revues sans dépôts (sur la période de travail 2010-2018)** du corpus ainsi que les caractéristiques essentielles de leur profil de dépôt sur Hal hors de la période de référence :

E_007	OpenEdition	S.A.P.I.E.N.S	5 dépôts réalisés entre 2009 et 2017 d'articles parus de 2008 à 2014, mais comprenant seulement des notices parus et déposés après 2010
G_036	CAIRN	Outre-Terre, Revue française de géopolitique	10 dépôts, tous effectués entre 2005 et 2014 pour des articles parus entre 2003 et 2008
EG_007	IDATE	DigiWorld Economic Journal	19 dépôts dans Hal dont seulement 2 articles parus en 2007
H_002	Papier	Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public	4 dépôts dénombrés dans Hal, donc 1 texte intégral correspondant à un article paru en 2003 et déposé en 2018
H_012	CAIRN	L'Année épigraphique ¹⁵	9 dépôts dans Hal tous sous forme de notices d'articles parus entre 2010 et 2018
H_014	CAIRN	Archives d'histoire doctrinale et	20 dépôts de notices d'articles parus entre 1994 et

¹⁵ *L'Année épigraphique* est un cas un peu particulier car ce titre n'est à proprement parler pas une revue, mais un ouvrage collectif de contributions non individualisées. Un outil de travail de type encyclopédique que véritablement une revue. Elle n'a de commun avec les revues son ISSN, sa périodicité annuelle, et sa présentation systématique et identique d'un volume à l'autre

		littéraire du Moyen Âge	2017 et déposés entre 2005 et 2018
H_017	JSTOR (1839-2013)	Bibliothèque de l'École des Chartes	1 notice déposée en 2018 pour un article paru en 2018
H_020	Persée (1941-2015)	Bulletin de la Société Nationale des Antiquaires de France	19 dépôts sur Hal, mais seulement sous formes de notices pour des articles parus de 1992 à 2018
H_025	Papier	Cahiers alsaciens d'archéologie d'art et d'histoire	21 dépôts dont 4 textes intégraux d'articles parus entre 1990 et 2001
H_031	CAIRN	Cahiers d'histoire de l'aluminium	Aucun dépôt sur Hal n'a été enregistré avant 2019, année où au 9 août on comptait 4 notices déposées
H_033	Papier	Cahiers d'Histoire et de Philosophie des Sciences	9 notices déposées et 1 article paru en 1992
H_036	OpenEdition & CAIRN (3 dernières années)	Cahiers du monde russe	43 notices et 4 articles déposés (parus entre 2000 et 2005)
H_040	Papier	Compostelle. Cahiers d'Études, de Recherche et d'Histoire Compostellanes	1 dépôt correspondant à un article en texte intégral paru en 2007 et déposé en 2010
H_072	Papier	Histoire et Archives	8 dépôts dont 7 notices et un texte intégral paru en 2005 et déposé en 2006
H_073	CAIRN	Histoire & sociétés rurales	6 notices déposées entre 2008 et 2012 pour des articles parus entre 2008 et 2011
H_074	Papier	Histoire et Sociétés : revue européenne d'histoire sociale	13 dépôts dont 12 notices et un texte intégral paru en 2007 et déposé la même année
H_075	Papier	Histoire médiévale et archéologie	2 dépôts de notices d'un article paru en 2006 et déposé en 2008 et d'un autre paru en 2017 et déposé en 2018
H_092	Revue1900	Mil neuf cent : Cahiers Georges Sorel : Revue d'histoire intellectuelle	12 dépôts dont 10 notices et 2 articles en texte intégral parus en 2004 et 2007 et déposés tous deux en 2010
H_096	CAIRN	Napoleonica. La revue	Aucun dépôt sur Hal
H_098	OpenEdition	Oliviana. Mouvements et dissidences spirituels XIIIe-XIVe siècles	4 dépôts de notices d'articles parus entre 2009 et 2016 et déposés entre 2009 et 2017
H_108	Peeters	Revue Biblique	8 notices déposées d'articles parus entre 1978 et 2014
H_109	OpenEdition	Revue d'Alsace	16 dépôts dont 9 notices et 7 articles en texte intégral parus entre 1997 et 2007 et déposés entre 2004 et 2017
H_110	CAIRN	Revue d'Assyriologie et d'Archéologie Orientale	10 dépôts dont 6 notices et 4 articles en texte intégral parus entre 1992 et 2000 et déposés entre 2011 et 2013
H_115	CAIRN	Revue des revues : revue internationale d'histoire et de bibliographie	1 dépôt de notice en 2014 pour un article paru en 2005
H_118	CAIRN	Revue d'histoire de la Shoah	20 dépôts dont 18 notices et 2 articles en texte intégral parus en 2004 et 2008 et déposés respectivement en 2018 et 2017
H_135	Papier	Sciences et Techniques en Perspective	16 dépôts dont 13 notices et 3 articles en texte intégral parus en 1998, 2003 et 2008 et déposés respectivement en 2005, 2018 et 2012
H_136	Papier	Sigila. Revue transdisciplinaire franco-portugaise sur le secret	11 dépôts dont 9 notices et 2 articles en texte intégral parus en 2004 et 2005 et déposés en 2005 et 2010
H_140	OpenEdition	Témoigner. Entre histoire et mémoire = Getuigen. Tussen Geschiedenis en Herinnering	10 notices déposées entre 2010 et 2017 pour des articles parus entre 2009 et 2017
H_145	OpenEdition	XVII-XVIII. Bulletin de la Société d'études anglo-américaines des	22 dépôts dont 16 notices et 6 articles en texte intégral parus entre 1991 et 2002 et déposés en

		XVIIe et XVIIIe siècles	2016 et 2017
L_002	CAIRN	Année balzacienne (L')	9 notices déposées d'articles parus entre 1998 et 2016 et déposées entre 2016 et 2018
L_005	Papier	L'Atelier du Roman	10 notices déposées d'articles parus entre 2008 et 2016 dont le dépôt s'est étalé entre 2012 et 2018
L_011	Papier	Bulletin de la Société Internationale des Etudes Yourcenariennes	2 notices déposées en 2011 pour des articles parus en 1989 et 1990
L_021	SAGE	Cahiers élisabéthains	20 notices déposées entre 2006 et 2016 pour des articles parus entre 1987 et 2016
L_025	Papier	Cahiers Valéry Larbaud	7 dépôts dont 5 notices et 2 textes intégraux parus en 2003 et 2008 et déposés tous deux en 2018
L_032	CAIRN	Etudes germaniques	43 dépôts dont 40 notices et 3 articles en texte intégral parus en 2002, 2006 et 2009 et respectivement déposés en 2014, 2016 et 2015
L_034	ERUDIT	Études littéraires africaines	35 dépôts dont 32 notices et 3 textes intégraux parus en 2007 et déposés entre 2008 et 2015
L_039	Papier	Histoires littéraires	18 dépôts dont 15 notices et 3 documents correspondant à des articles parus en 2005, 2006 et 2007 et déposés respectivement en 2009, 2014 et 2014
L_041	Papier	Infini	5 notices déposées entre 2006 et 2016 pour des articles parus entre 2003 et 2012
L_045	Icar.cnrs.fr	Langues et littératures du monde arabe	28 dépôts dont 25 notices et 3 articles en texte intégral parus en 2003, 2006 et 2007 et déposés respectivement en 2017, 2009 et 2009.
L_052	Papier	Nouvelle Revue Française	15 dépôts dont 14 notices et 1 article en texte intégral paru en 1999 et déposé en 2017
L_053	Papier	Otrante. Arts et littératures fantastiques	33 dépôts dont 32 notices et 1 article en texte intégral paru en 2009 et déposé en 2017
L_055	CAIRN	Po&sie	6 notices déposées en 2010 et en 2017 pour des articles parus entre 1990 et 2016
L_057	CAIRN	Présence Africaine	3 dépôts dont 2 notices et 1 article en texte intégral paru en 2000 et déposé en 2008
L_063	Papier	Revue d'études proustiennes	2 notices déposées en 2018 pour des articles parus en 2016
L_065	CAIRN	Revue de philologie, de littérature et d'histoire anciennes	19 dépôts dont 18 notices et 1 texte intégral d'un article paru en 2000 et déposé en 2017
L_080	Papier	Siècle 21. Littérature et société	9 notices déposées d'articles parus entre 2007 et 2016
L_086	ubu-apite.org	Ubu	2 notices déposées en 2012 pour des articles parus en 1997 et 1999

Tableau 8 : Liste des revues ne présentant aucun dépôt d'articles parus entre 2010 et 2018

III.1.5 Qui sont les déposants ?

L'implication des auteurs dans le dépôt de leurs articles

Hal dispose d'une métadonnée permettant d'identifier l'auteur du dépôt de la notice ou du document : il s'agit de la métadonnée « Nom complet du contributeur/déposant », qu'il est possible d'ajouter aux champs à exporter pour informer les dépôts correspondant à une revue. **On peut ainsi mesurer la fréquence de la non-coïncidence entre « nom d'auteur » et « nom du contributeur/déposant », qui donne une information précieuse sur le rôle de tiers-déposant. Le rôle des laboratoires, des bibliothèques universitaires et autres services d'information scientifique et technique (IST) dans le dépôt sur Hal peut être appréhendé tout comme l'éventuelle**

contribution de tierces personnes ne faisant pas partie des coauteurs (on peut penser à des collègues de laboratoire, par exemple, même s'il s'agit d'une hypothèse à tester).

Quoi qu'il en soit, voici les résultats sur l'intégralité du corpus : **sur les 27 162 documents déposés, notices et textes intégraux confondus, 10 361 l'ont été par l'auteur ou l'un des co-auteurs de l'articles, ce qui représente environ 38% des dépôts. Si l'on s'intéresse au seul dépôt d'articles en texte intégral, qui représente 6889 documents, ce pourcentage augmente de façon significative puisqu'il atteint 54% du sous-corpus, avec 3695 articles. Les auteurs sont donc plus engagés dans le dépôt des textes intégraux de leurs articles que dans celui de notices descriptives.**

Qu'en est-il de la répartition de ces chiffres par discipline ? Le tableau suivant donne les statistiques établies précédemment pour les 5 disciplines et depuis la mise en service de Hal en 2002 :

Disciplines	Taux d'articles en texte intégral déposés par l'un des co-auteurs
Environnement	147/254 ≈ 58%
Eco-gestion	1122/2014 ≈ 56%
Géographie	1737/2244 ≈ 77%
Histoire	635/1700 ≈ 37%
Littérature	54/676 ≈ 8%

Tableau 9 : Implication des auteurs dans le dépôt de leurs articles

Les différences disciplinaires sont très marquées. La géographie et la littérature illustrent les comportements les plus opposés des auteurs : en géographie, plus de 3 articles sur 4 (77%) se trouve déposé par son auteur ou l'un de ses co-auteurs dans Hal, tandis que c'est moins d'1 article sur 10 qui l'est en littérature (8%). Les trois autres disciplines ont des comportements plus moyens, l'histoire étant plutôt dans la moyenne basse (37% des auteurs impliqués dans le dépôt) et l'éco-gestion et l'environnement dans la moyenne haute (56% et 58% respectivement).

On pourrait penser que les écarts importants constatés sont dus simplement à la structure de l'autorat des articles, plus collective en économie-gestion ou en sciences de l'environnement qu'en histoire ou en littérature. Or, les statistiques montrent que s'il y a effectivement des différences disciplinaires notables en termes de nombre d'auteurs par article, elles ne suffisent pas à elles seules à expliquer les écarts très importants constatés en matière de comportement des déposants. En effet, les statistiques sur la base de données complète des dépôts dans Hal montrent que pour la littérature et l'histoire, on compte respectivement 1,1 et 1,3 auteurs en moyenne par article, tandis que pour la géographie et l'éco-gestion les articles sont en moyenne écrits à 2 auteurs (1,97 pour la géographie et 1,99 pour l'éco-gestion), la discipline à l'autorat le plus collectif étant sans surprise les sciences de l'environnement¹⁶, où l'on compte 2,4 auteurs par article en moyenne.

A grands traits, on peut donc dire qu'il y a deux fois plus d'auteurs par articles en sciences de l'environnement, géographie et éco-gestion qu'en histoire et en littérature, ce qui explique peut-être les différences entre le taux d'articles déposés par leur auteur en histoire (37%) et celui en

¹⁶ Sans surprise car les sciences de l'environnement sont la discipline du corpus dans laquelle on trouve la plus grande proportion de revues de sciences naturelles, sciences dans lesquelles la structure de la recherche de laboratoire et partant de la rédaction d'articles, est plus collective qu'en sciences humaines et sociales.

géographie (avec 77%, on s'approche effectivement d'un taux deux fois plus élevé), mais qui ne rend pas compte de l'écart entre la littérature et les autres disciplines.

Qui sont les tiers déposants ?

Pour une première approche, on travaille sur le sous-corpus des textes intégraux déposés dans Hal depuis 2002. Ce sont en tout 3194 documents de ce type qui ont été déposés par des tiers dans les revues des cinq disciplines d'étude (sur les 6889 articles en texte intégral du corpus). A l'aide d'un tableau croisé dynamique, on établit le nombre de déposants associés à chaque nombre de dépôts, et on ordonne le résultat par ordre décroissant, ce qui permet de faire ressortir les plus gros déposants. Le résultat est le suivant pour les 30 premiers tiers-dépôts du corpus :

Statut du tiers-déposant	Nombre de dépôts effectués
Archive ouverte Spire de Sciences Po	568
Ingénieur d'études chargée de ressources documentaires LAET (UMR 5593), Lyon	84
Archive ouverte de l'IRSTEA	74
Administrateur HAL de Nanterre	61
Correspondante IST pour le CESCO (UMR 7302)	55
Assistante administrative (IREGE, Institut de Recherche en Gestion et en Economie)	52
Cadic de l'IFSTTAR	51
Ingénieur d'études du Centre Interlangues (EA 4182), Université de Bourgogne, Dijon	49
Chargé de la stratégie numérique de l'IFSTTAR (Direction Scientifique)	46
Responsable du Centre de Documentation du laboratoire GEODE (CNRS, Toulouse 2)	41
Ingénieur d'études CNRS documentaliste du Laboratoire d'Economie Appliquée de Grenoble (GAEL)	32
Archive ouverte de l'INRA (Prodinra)	31
Bibliothécaire chez Audencia Business School (Nantes)	30
Assistante de communication du laboratoire ARTEHIS : Archéologie, Terre, Histoire, Sociétés (UMR 6298), Université de Bourgogne, Dijon	29
Secrétaire de rédaction de la revue Review of Agricultural, Food and Environmental Studies (INRA)	29
Directrice adjointe de la bibliothèque de l'Université Paris-Dauphine-PSL	27
Assistante ingénieur (CNRS) du laboratoire ARCHIMEDE (Archéologie et Histoire Ancienne : Méditerranée-Europe) (UMR 7044), Strasbourg	26
Documentaliste du Centre Georges Chevrier (Sociétés et Sensibilités) (UMR 7366), Dijon, Université de Bourgogne	23
Ingénieur d'étude documentaliste du Groupe d'Analyse et de Théorie Economique (GATE) (UMR 5824), Lyon-St-Etienne	23
Ingénieur d'études CNRS du centre de recherche PROTASI (Proche Orient, Turquie, Asie Intérieure) - UMR 7571	23
Compte Hal de l'UMR Prodig	21
Responsable du service Production scientifique de la Bibliothèque Universitaire Centrale de l'Université de Rennes 2	17
Compte Hal de l'Ecole Française d'Extrême-Orient	17
Professeur d'Université en Langues et Littérature françaises des XXème et XXIème siècles à l'Université de Lille,	15
Compte Hal de l'Université de la Réunion	13
Chargée de ressources documentaires CNRS, Laboratoire AAU (Ambiances, Architectures, Urbanités), Grenoble	12
Compte Hal du Centre Georges Chevrier, Université de Bourgogne	12
Professeur d'Université en histoire médiévale à l'Université Lyon 2, codirectrice de la revue Médiévales	11
Archiviste-Documentaliste à Agrocampus Ouest	11
Documentaliste et Webmestre du LA3M (Laboratoire d'Archéologie Médiévale et Moderne en Méditerranée), UMR 7298, Aix-en-Provence	10

Tableau 10 : Statut des 30 premiers tiers-déposants classés par ordre décroissant de nombre de dépôts de textes intégraux

Sur les 30 premiers tiers-déposants, on peut déjà identifier, simplement à partir du nom inscrit dans Hal, 9 archives ouvertes ou services tiers d'université et de laboratoires :

- **Les archives ouvertes**¹⁷ : Spire Sciences Po, CEMOA de l'IRSTEA (Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture), ProdINRA de l'INRA
- **Les services centraux de la documentation des universités et établissements d'ESR** : l'administrateur Hal de l'Université Paris-Ouest Nanterre, l'IFFSTAR CADIC, l'Ecole Française d'Extrême-Orient (EFEO), Université de la Réunion.
- **Les laboratoires** : UMR PRODIG (CNRS, Paris-Diderot, Paris-Sorbonne, etc.), Centre Georges Chevrier (Sociétés et Sensibilités) de l'Université de Bourgogne.

Une rapide recherche sur Google permet de renseigner le statut professionnel des autres personnes tiers-déposantes : 18 d'entre elles sont **des professionnelles de l'IST** (ce ne sont effectivement que des femmes), 2 sont **des professeurs d'université** (un homme en littérature et une femme en histoire médiévale) et on trouve enfin l'ancien directeur de la stratégie informatique de l'IFSTTAR.

Bien entendu, il s'agit là des tiers-déposants les plus actifs, et l'on ne peut extrapoler ces résultats à l'intégralité du corpus, qui est fait pour la plupart des dépôts de contributeurs uniques, comme le montre le graphique ci-dessous :

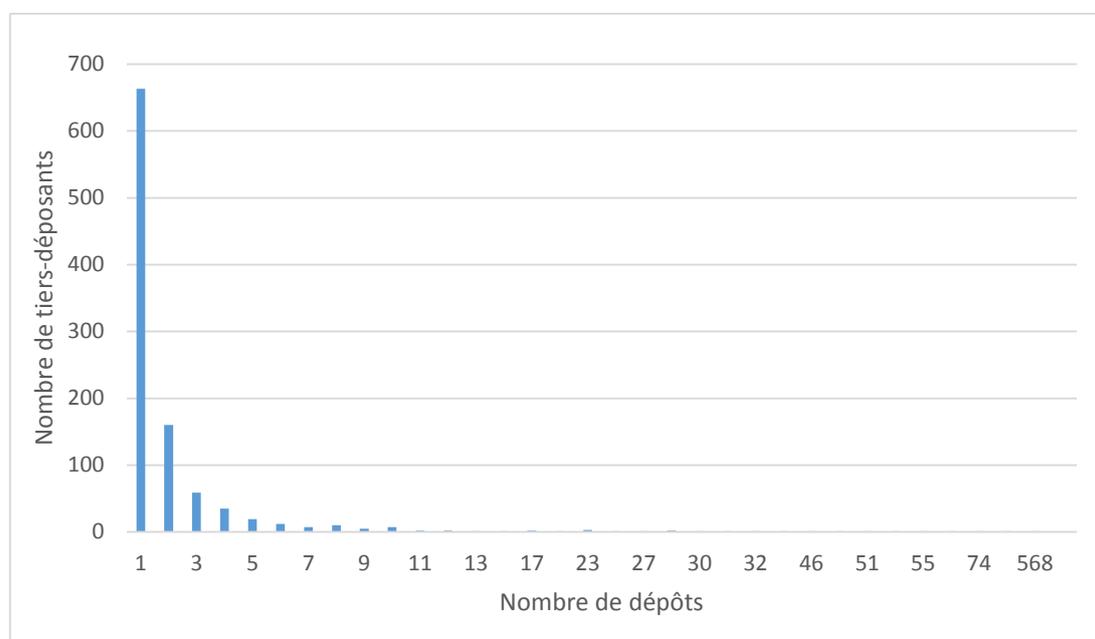


Figure 4 : Nombre de tiers-déposants selon le nombre de dépôts dans Hal (textes intégraux, 2002-2018)

Ces données sur les « contributeurs intensifs » **devront donc être complétés par une étude du statut professionnel des tiers-déposants réalisée sur un échantillon représentatif de la totalité du corpus.**

¹⁷ Notons que l'on ne peut pas savoir si l'auteur du dépôt dans une tierce archive institutionnelle est un auteur ou un tiers, et qu'il s'agit donc de tiers dépôts spécifiques par rapport à ceux réalisés par les services tiers.

III.2 Analyse dynamique des dépôts dans Hal

III.2.1 Evolution des dépôts sur la période 2010-2018

On peut représenter graphiquement l'évolution des dépôts dans Hal sur la période de travail pour les cinq disciplines concernées :

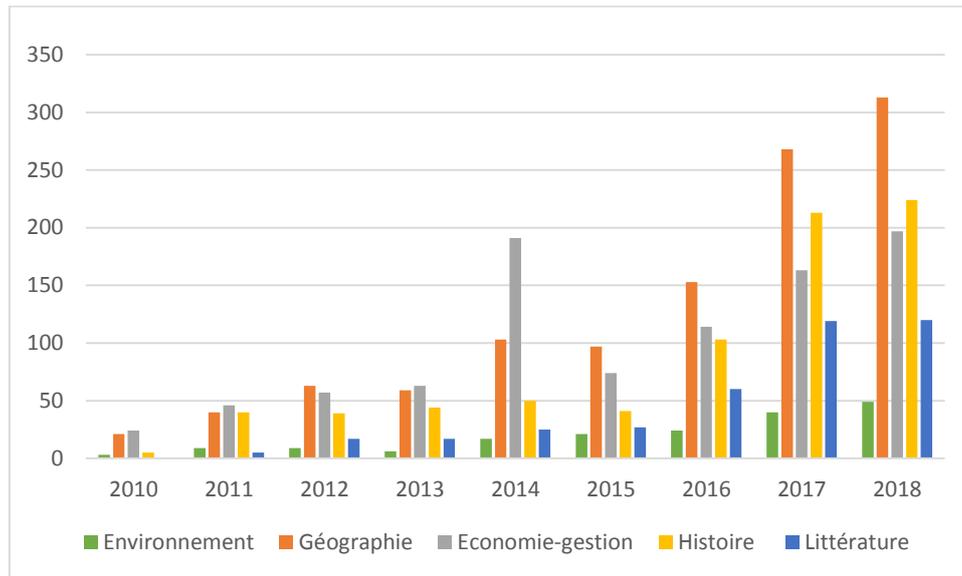


Figure 5 : Evolution du nombre de dépôts annuels dans Hal par discipline pour les articles parus de 2010 à 2018

Ce premier aperçu de l'évolution du nombre de dépôts de 2010 à 2018 doit être pris avec précaution car il ne fait figurer que les dépôts d'articles ayant paru de 2010 à 2018. Ainsi, pour l'année 2010, le stock d'articles parus et susceptibles d'être déposés est très faible tandis que pour l'année 2018 il a de fortes chances d'être beaucoup plus grand car il se compose de tous les articles parus entre 2010 et 2018 et n'ayant pas encore été déposés.

Pour corriger ce biais d'un regard porté simplement sur le nombre de dépôts par an, on peut établir le ratio entre le nombre de dépôts effectué chaque année et le stock d'articles à déposer disponible cette année-là. Soit $N_{\text{parus}}(\text{Année } 1)$ le nombre d'articles parus en T1, soit $N_{\text{déposés}}(\text{Année } 1)$ le nombre d'articles déposés en T1, le stock d'articles disponibles au dépôt en T2, noté $N_{\text{stock}}(\text{T2})$ est alors :

$$N_{\text{stock}}(\text{Année } 2) = [N_{\text{parus}}(\text{Année } 1) + N_{\text{parus}}(\text{Année } 2)] - N_{\text{déposés}}(\text{Année } 1)$$

De même, pour l'année 3 :

$$N_{\text{stock}}(\text{Année } 3) = [N_{\text{parus}}(\text{Année } 1) + N_{\text{parus}}(\text{Année } 2) + N_{\text{parus}}(\text{Année } 3)] - [N_{\text{déposés}}(\text{Année } 1) + N_{\text{déposés}}(\text{Année } 2)]$$

On a donc :

$$N_{\text{stock}}(\text{Année } 3) = N_{\text{stock}}(\text{Année } 2) + N_{\text{parus}}(\text{Année } 3) - N_{\text{déposés}}(\text{Année } 2)$$

On peut alors calculer pour chaque revue et pour chaque année un nombre d'articles en stock pour le dépôt. Le tableau ci-dessous donne les valeurs moyennes de ce nombre pour chaque discipline :

	2010	2011	2012	2013	2014	2015	2016	2017
N _{stock} (Env.)	23,92	49,46	79,61	107,69	78,38	107,54	79,23	110,92
N _{stock} (Géo)	26,88	45,57	68,90	88,72	111,61	131,71	150,66	169,47
N _{stock} (Eco-gest.)	23,31	44,39	70,37	93,91	117,31	137	157,78	176,96
N _{stock} (Histoire)	20,935	34,297	50,95	70,29	90,39	108,97	128,05	145,08
N _{stock} (Litt.)	17,21	35,83	56,98	77,85	97,90	119,66	141,19	163

En faisant figurer ces valeurs dans le graphique précédent, on mesure mieux le risque qu'il y a à considérer l'évolution du nombre de dépôts sans tenir compte de l'évolution parallèle du stocks d'articles disponibles pour le dépôt :

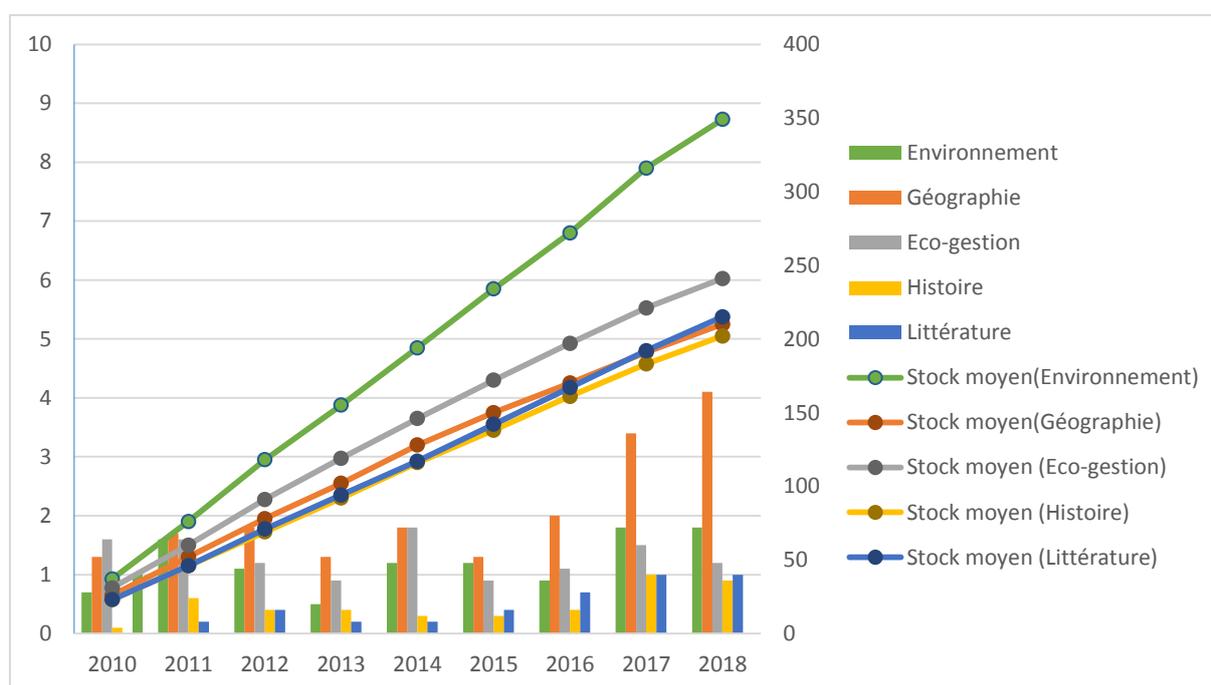


Figure 6 : Evolution parallèle de l'intensité dépositrice dans Hal (en %, sur l'axe de gauche) et du stock d'articles disponibles au dépôt pour cinq disciplines (en nombre d'articles, sur l'axe de droite).

Indication de lecture : L'axe des ordonnées de gauche donne le pourcentage de dépôts dans Hal par rapport au stock d'articles disponibles, stock qui est justement représenté sur l'axe des ordonnées de droite donne le stock moyen d'articles disponibles par discipline.

La proximité des barres des histogrammes donnant le nombre de dépôts annuels dans Hal et des droites représentant les stocks d'articles disponibles au dépôt ne doit pas faire illusion : l'ordre de grandeur des dépôts dans Hal est l'unité, celui du nombre d'articles en stock est la centaine (quelques dizaines d'abord, puis la centaine dès 2013-2014). En effet, c'est toujours moins de 5% du stock d'articles disponibles que l'on retrouve en texte intégral sur Hal, ce qui veut dire que dans 95% des cas, les articles parus sur la période 2010-2018 ne se retrouvent pas sur Hal.

Pour analyser la situation plus finement, on peut s'intéresser à l'indicateur qui rapporte le nombre de dépôts effectués chaque année par l'ensemble des revues au nombre d'articles disponibles en stock année après année, et suivre l'évolution de cet indicateur. Comme précédemment, on reporte les données agrégées par disciplines du rapport nombre de dépôts effectués à l'année A/nombre d'articles disponibles au dépôt à l'année A (stock) dans un tableau de synthèse :

Intensité déposante discipline	par	2010	2011	2012	2013	2014	2015	2016	2017	2018
Environnement		0,7	1,6	1,1	0,5	1,2	1,2	0,9	1,8	1,8
Géographie		1,3	1,7	1,8	1,3	1,8	1,3	2	3,4	4,1
Eco-gestion		1,6	1,6	1,2	0,9	1,8	0,9	1,1	1,5	1,2
Histoire		0,1	0,6	0,4	0,4	0,3	0,3	0,4	1	0,9
Littérature		0	0,2	0,4	0,2	0,2	0,4	0,7	1	1
Toutes disciplines		0,6	0,9	0,8	0,6	0,8	0,6	0,9	1,5	1,6

Tableau 11: Evolution de l'intensité déposante par discipline sur la période de travail (2010-2018)

A la lecture de ce tableau, on voit d'abord que la progression linéaire que l'on constatait pour le nombre de dépôts par an pour chaque discipline disparaît lorsqu'on rapporte ce nombre de dépôts au stock d'articles disponibles chaque année. Ainsi, pour l'environnement, la géographie et l'éco-gestion, après une légère hausse de la valeur de ce rapport en 2011, on passe par une baisse répétée en 2012 et en 2013. La seule hausse notable de l'« intensité déposante » des revues que l'on constate est celle qui marque l'année 2017. En effet, cette année-là, toutes les disciplines connaissent une forte hausse de l'intensité déposante : multiplication par presque 2 de ce pourcentage pour la géographie, l'histoire et les sciences de l'environnement et par 1,5 pour l'éco-gestion et la littérature. L'année 2018, en revanche, représente une stagnation de l'intensité déposante pour toutes les disciplines à l'exception de la géographie. Une hypothèse possible est qu'après un essor rapide des dépôts suite à la loi pour la République numérique de 2016 et aux initiatives et débats que celle-ci a suscité, on entrerait dans un régime de croisière des dépôts sur archives ouvertes, hypothèse que seules des mesures pour les années futures permettront d'étayer.

Ces chiffres ne doivent pas pour autant faire oublier que les pourcentages en question sont très faibles : ainsi, pour la littérature et l'histoire, ce sont moins de 1% des articles du stock disponible au dépôt qui sont chaque année déposés sur Hal. Ces pourcentages sont un peu plus élevés pour la géographie et l'éco-gestion mais ils restent faibles néanmoins.

III.2.2 Définition et évolution du taux de dépôt des articles parus dans l'année

Nous voudrions maintenant **éclairer les pratiques de dépôt rapide**, c'est-à-dire suivant de peu la publication de l'article dans la revue l'ayant accepté. La Loi pour la République numérique d'octobre 2016 autorise les chercheurs en SHS ayant disposé d'un financement public équivalent à la moitié au moins de leur budget à auto-diffuser leurs publications au bout d'un an d'embargo. Cet indicateur, comparé à celui de dépôt au bout d'un an et avant la période de fin d'embargo, permet de mesurer finement l'impact de la loi sur les pratiques de dépôt des chercheurs et des services tiers accompagnant la recherche¹⁸.

¹⁸ Une limite de cet indicateur est qu'il se base sur la comparaison des dates de parution et des dates de libération du dépôt dans Hal, à l'échelle de l'année. C'est-à-dire que si un article est paru en décembre 2014 et déposé dans Hal en novembre 2015, il ne sera pas compté par notre calcul comme ayant été déposé dans

	En %	2010	2011	2012	2013	2014	2015	2016	2017
$N_{\text{dépôts}}/N_{\text{parutions annuelles}}$ (Env.)		0,322	1,201	1,496	0,267	0,777	2,771	1,282	2,733
$N_{\text{dépôts}}/N_{\text{parutions annuelles}}$ (Géo)		3,788	0,761	0,948	1,46	1,991	2,644	4,85	2,934
$N_{\text{dépôts}}/N_{\text{parutions annuelles}}$ (Ecogest)		2,273	2,482	2,54	1,901	1,095	0,743	3,846	4,461
$N_{\text{dépôts}}/N_{\text{parutions annuelles}}$ (Histoire)		0,231	0,204	0,212	0,255	0,243	0	0,473	0,639
$N_{\text{dépôts}}/N_{\text{stock annuel}}$ (Litt.)		0	0,19	0	0,152	0	0,148	0,449	1,362

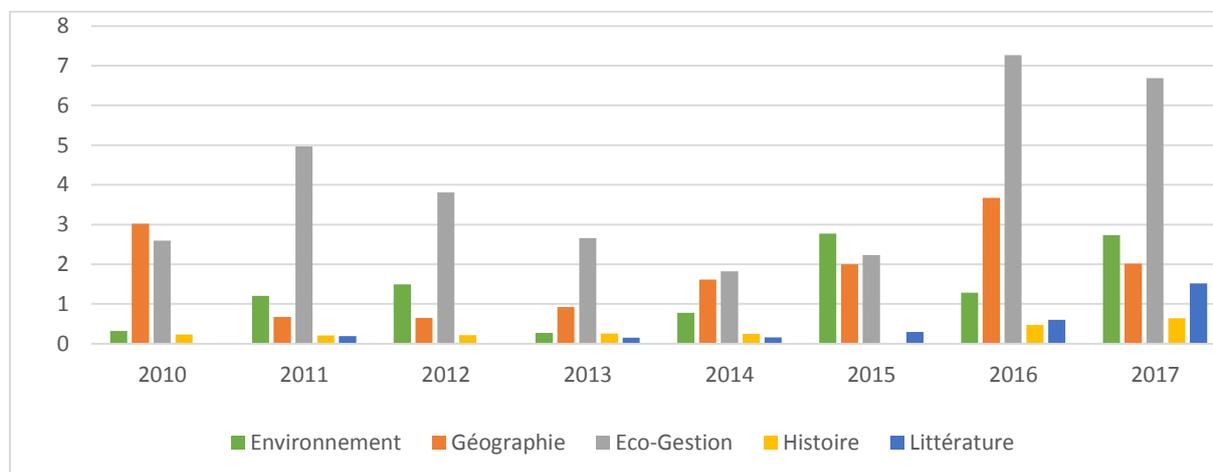


Figure 7 : Evolution du taux de dépôts sur Hal dans l'année

La progression régulière constatée précédemment pour l'intensité dépositrice entre 2015 et 2017 n'est pas aussi claire ici. En réalité, elle ne concerne que la littérature et l'histoire. En revanche, ce qu'on peut dire, c'est **que les trois dernières années présentent des taux de dépôt dans l'année bien plus élevés que ceux des années précédentes dans toutes les disciplines**. Globalement, il semblerait que la tendance soit à une accélération de la vitesse de libération des articles sur Hal au fur et à mesure qu'on progresse dans le temps.

III.3 Délai de dépôt en archive ouverte des articles dans Hal et barrière mobile

Afin de comprendre la dynamique temporelle du dépôt sur l'archive ouverte Hal et en particulier le rôle joué par les barrières mobiles appliquées par de nombreuses revues, on peut définir un **délai de dépôt en archive ouverte d'un article comme l'intervalle de temps entre sa parution et sa mise en ligne sur Hal en texte intégral**.

III.3.1 Résultats au niveau de la revue

On peut calculer cet indicateur grâce au renseignement des deux dates dans la base de données de Hal (ce qui n'est pas possible sur Academia). Une fois calculé le délai de dépôt en archive ouverte moyen des articles en textes intégraux par revues, on fait la moyenne par discipline et on obtient un indicateur du délai de dépôt en archive ouverte des articles par discipline. Notons que la moyenne

l'année suivant la parution, ce qui aurait dû être le cas. Les chiffres que nous proposons sous-estiment donc le nombre réel de dépôts dans l'année.

est faite en attribuant la même pondération à chaque revue, ce qui a pour effet de mettre sur un pied d'égalité des revues pour lesquels un texte intégral a été déposé sur Hal avec des revues qui en auraient déposé plusieurs dizaines.

	Environnement	Géographie	Eco-gestion	Histoire	Littérature
Délai de dépôt en archive ouverte moyen des articles par discipline (en mois)	25	29	25	36	30
Barrière mobile moyenne par discipline (en mois)	17	24	23	34	15

Tableau 12 : Délai de dépôt en archive ouverte moyen et barrière mobile moyenne par discipline

On voit que **le délai de dépôt en archive ouverte d'un article n'est jamais inférieur à deux ans**, même lorsque les barrières mobiles sont inférieures en moyenne, à deux années. On note ensuite des comportements contrastés selon les disciplines : **alors que pour l'histoire, le délai de dépôt en archive ouverte moyen est à peu près équivalent à la barrière mobile moyenne, pour la littérature, il est deux fois plus long que la barrière mobile**, qui est par ailleurs plus courte étant donné les nombreuses revues en accès ouvert immédiat.

III.3.2 Résultats au niveau détaillé de l'article

En complétant la base de données de dépôts dans Hal rassemblant tous les articles et notices déposés depuis 2002 pour les 368 revues du corpus avec les données de barrière mobile extraites sur Mirabel, on peut, en comparant la date de mise en ligne de l'article et la date de levée de barrière mobile après parution, **mesurer le comportement des déposants en matière de prise en compte de la barrière mobile et l'évolution de celui-ci avec le temps.**

Concrètement, il suffit de **comparer le délai de dépôt en archive ouverte de l'article sur Hal et la barrière mobile**, et si le premier est inférieur à la seconde, alors le dépôt du texte intégral a été fait avant la levée de la barrière mobile.

Le tableau suivant donne résultats globaux pour les 157 revues du corpus disposant d'une barrière mobile (on a donc éliminé celles en accès payant¹⁹ et celles en accès ouvert). Cela représente 3420 articles en texte intégral déposés de 2002 à 2018 et 1423 pour la période 2010-2018 :

¹⁹ Revues dont l'accès aux articles est payant sur toute la période de travail.

	Nombre et proportion des dépôts réalisés avant la levée de la barrière mobile
Intégralité du corpus (3420 articles, 2002-2018)	1160 (25,3%)
Corpus récent (1423 articles, 2010-2018)	847 (37,3%)

Tableau 13 : Nombre et proportion des dépôts réalisés avant la levée de la barrière mobile

Sur la période récente, on constate une accélération de la vitesse de dépôt qui a pour conséquence que les dépôts sont de plus en plus effectués avant la levée de la barrière mobile sur la version éditeur de l'article.

Qu'en est-il de la répartition par discipline de ces comportements ? Le tableau suivant répond à la question²⁰ :

	Nombre et proportion des dépôts réalisés avant la levée de la barrière mobile (2002-2018)	Nombre et proportion des dépôts réalisés avant la levée de la barrière mobile (2010-2018)
Environnement	74 (29,1%)	64 (35,9%)
Eco-gestion	440 (21,9%)	318 (34,2%)
Géographie	361 (16,1%)	243 (21,7%)
Histoire	229 (14,2%)	171 (22,5%)
Littérature	56 (8,2%)	51 (13,1%)

Tableau 14 : Nombre et proportion des dépôts réalisés avant la levée de la barrière mobile pour les deux périodes 2002-2018 et 2010-2018

Sur la période récente, on a ainsi environ 1/3 des articles déposés avant la levée de la barrière mobile en environnement et en éco-gestion, taux qui descend à environ 1/5 pour la géographie et l'histoire et à 1/10ème pour la littérature.

Barrière mobile et taux de dépôt

Afin de mesurer l'influence de la barrière mobile sur le taux de dépôt des articles dans Hal, on peut réaliser un test de Student entre la variable binaire « existence ou non d'une barrière mobile » et la variable quantitative à expliquer « taux de dépôts dans Hal ».

Sur les trois premières disciplines du corpus (à l'exclusion de l'histoire et de la littérature), on obtient les résultats suivants avec le logiciel R. Le test est bien significatif puisque la p-valeur est de 0,01525 et le test montre que **l'existence d'une barrière mobile ou d'un accès payant diminue en moyenne le taux de dépôts de 6%, c'est-à-dire qu'elle le divise par 2 puisque le taux de dépôts des revues sans barrière mobile est de 13,8%.**

²⁰ Attention, dans le tableau ci-dessous, les pourcentages sont calculés par rapport à l'intégralité du corpus d'articles déposés par discipline et non seulement par rapport aux articles publiés dans les revues à barrière mobile.

IV. Etude de la diffusion des articles de recherche en SHS sur les réseaux sociaux académiques Academia et ResearchGate

Pour obtenir des résultats comparables à celles détenues sur Hal sur Academia et ResearchGate, on a employé la méthode suivante :

- Constitution de bases de données d’articles Titre-Auteurs-Dates de mise en ligne pour la période de travail : 2010-2018
 - Open Edition : par extraction OAI-PMH
 - CAIRN : données livrées par la plateforme
- Rédaction d’un script qui interroge le moteur de recherche d’Academia à partir de la base de données Titre-Auteurs-Dates et renvoie l’information de la présence ou de l’absence du titre de l’article dans la base d’Academia²¹.
- Vérification manuelle de la pertinence des articles obtenus : s’agit-il bien d’articles de recherche et pas de notes de lectures, articles d’actualité, etc. ?
- Même méthode sur ResearchGate mais nécessité d’un échantillonnage aléatoire à cause du serveur de ResearchGate qui rejette le script au bout d’un certain nombre de requêtes.

Cela signifie que **le corpus de travail pour les analyses concernant les réseaux sociaux académiques se restreint au sous-corpus CAIRN+OpenEdition (166 revues)**, et que nous n’avons pas pu prendre en compte les revues en dehors de ces deux plateformes étant donné les difficultés pour établir à la main les bases de données d’articles parus de 2010 à 2018.

Sur ce sous-corpus, voici les résultats agrégés :

Données obtenues sur le sous-corpus CAIRN + OpenEdition	Moyenne du nombre de dépôts par an pour une revue (2010-2018)		
	HAL (1682 articles en texte intégral)	Academia (2485 articles en texte intégral)	ResearchGate
Environnement (8 revues)	1,3	2	0,2
Géographie (19 revues)	1,6	2,4	0,4
Eco-gestion (53 revues)	1,6	1,1	0,3
Histoire (80 revues)	0,8	1,8	< 0,1
Littérature (13 revues)	0,8	1,3	< 0,1

Tableau 15 : Nombre de dépôts moyen par revue réalisés sur Hal, Academia et ResearchGate sur la période 2010-2018 ;

²¹ Dans la mesure bien entendu où la graphie est le même que celle consignée dans nos bases de données. Il y a donc une proportion de faux négatifs que nous n’avons pas estimée et qui tient à ces différences de graphie.

Indication de lecture : en moyenne, sur la période 2010-2018, on compte 0,8 dépôts dans Hal par an pour une revue d'histoire, environ 1,8 sur Academia et moins de 0,1 sur ResearchGate²².

Ainsi, Academia semble présenter plus de textes intégraux que Hal pour 4 des 5 disciplines du corpus : environnement, géographie, histoire et littérature. Seule l'éco-gestion fait exception. C'est pour l'histoire que la différence est la plus marquée, avec plus de deux fois plus de dépôts sur Academia que sur Hal. Pour la littérature, la géographie et les sciences de l'environnement, ces taux sont plus proches de 1,5.

Les dépôts sur Hal et Academia sont-ils les mêmes ?

En mesurant la part des doublons existant entre les dépôts effectués sur Hal et Academia, on peut déterminer si ce sont les mêmes articles qui sont à la fois déposés sur Hal et sur Academia ou bien si ce sont des articles différents. Dans la perspective d'établir un taux d'articles disponibles en accès gratuit sur Internet (qui prend en compte les articles disponibles en accès ouvert sur les plateformes des éditeurs), cette mesure est importante.

Sur les 2485 articles en texte intégral présents sur Academia relevant de l'une des 169 revues du sous-corpus CAIRN + OpenEdition, combien en trouve-t-on parmi les dépôts dans Hal ? Le tableau suivant résume les résultats obtenus sur la période 2010-2018, et en prenant en compte également les articles présents sur Academia et existant seulement comme notices sur Hal.

	Articles présents en texte intégral sur Hal et Academia	Articles présents en texte intégral sur Academia et sous forme de notice sur Hal
Période de travail 2010-2018	92	202

Tableau 16 : Nombre de doublons Hal/Academia

Le sous-corpus CAIRN + OpenEdition représente environ 52 000 articles parus de 2010 à 2018. Si l'on somme les dépôts de textes intégraux sur Hal avec les articles trouvés sur Academia et que l'on retranche les 92 doublons, on a la totalité des articles en texte intégral disponible en accès gratuit soit sur Hal soit sur Academia et l'on peut calculer la part du total d'articles que cela représente.

Total des articles en accès gratuit sur Hal et Academia

=

$$1682 + 2485 - 92 = 4075$$

Cela représente environ 8 % des 52 000 articles parus de 2010 à 2018 sur le sous-corpus CAIRN + OpenEdition. On ne peut pas facilement extrapoler ce résultat à l'intégralité du corpus de revues, car les revues diffusées sur OpenEdition ou CAIRN ont plus de chance de trouver leurs articles disponibles sur Internet que les autres revues du corpus (ne serait-ce que parce qu'on compte parmi elles toutes les revues existant seulement au format papier), mais on peut estimer que l'ordre de grandeur de 5 à 8% d'articles en texte intégral disponibles en dehors des plateformes et sites de diffusion des éditeurs semble vraisemblable.

²² Pour obtenir ces moyennes, on recense le nombre de dépôts dans Hal, Academia et ResearchGate, que l'on divise ensuite par le nombre de revues dans la discipline concernée et par le nombre d'années de la période de travail, à savoir 9 (2010-2018).

Bibliographie

Mahé Annaïg et Prime-Claverie Camille, « Qui dépose quoi sur Hal-SHS ? Pratiques de dépôts en libre accès en sciences humaines et sociales », Revue française des sciences de l'information et de la communication, 11, 2017, <https://journals.openedition.org/rfsic/3315>

Mahé Annaïg et Prime-Claverie Camille, « Participation, contribution de l'auteur dans les archives ouvertes. D'une position militante à la prescription institutionnelle », Les Cahiers de la SFIC, 2014, p. 191-193.

Hornik, 2017, « Metadata harvesting with R and OAI-PMH ».

DGMIC/SLL, Mai 2018, Étude sur l'économie des revues de SHS.

Annexes

Annexe 1 : Liste des revues, précédées de leur numéro dans le corpus général et de leur identifiant disciplinaire

1	EG_001	@GRH	49	EG_049	Revue de l'Organisation Responsable
2	EG_002	Actes de la Recherche en Sciences Sociales	50	EG_050	Revue de la Régulation
3	EG_003	Annales d'Economie et de Statistique	51	EG_051	Revue de Philosophie Economique / Review of Economic Philosophy
4	EG_004	Bulletin Français d'Actuariat	52	EG_052	Revue Economique
5	EG_005	Cahiers d'Economie Politique	53	EG_053	Revue Française d'Economie
6	EG_006	Cahiers scientifiques du transport	54	EG_054	Revue Française de Gestion
7	EG_007	DigiWorld Economic Journal (anciennement Communications & Stratégies)	55	EG_055	Revue Française de Gouvernance d'Entreprise
8	EG_008	Comptabilité, Contrôle, Audit	56	EG_056	Revue Française d'administration Publique
9	EG_009	Concurrences	57	EG_057	Revue Française de Socio-économie
10	EG_010	Décisions Marketing	58	EG_058	Revue Française du Marketing
11	EG_011	Economie et Institutions	59	EG_059	Revue Interdisciplinaire sur le Management et l'Humanisme (RIMHE)
12	EG_012	Economie et Prévision	60	EG_060	Revue Internationale de Droit Economique
13	EG_013	Économie et Statistique	61	EG_061	Revue Internationale de l'Economie Sociale - RECMA
14	EG_014	Économie Publique	62	EG_062	Revue Internationale de Psychosociologie et de Gestion des Comportements Organisationnels
15	EG_015	Economie Rurale	63	EG_063	Revue internationale des études du développement (anciennement Revue Tiers Monde, avant 2017)
16	EG_016	Entreprendre & Innover	64	EG_064	Sciences Sociales et Santé
17	EG_017	Finance	65	EG_065	Systèmes d'Information et Management
18	EG_018	Finance Contrôle Stratégie	66	EG_066	Travail et Emploi
19	EG_019	Gérer et Comprendre			
20	EG_020	Gestion 2000	67	G_001	Annales de géographie
21	EG_021	Gestion et Management Public	68	G_002	Annales de la recherche urbaine
22	EG_022	Innovations - Revue d'économie et de management de l'innovation	69	G_003	ArchéoSciences
23	EG_023	Journal de Gestion et d'Economie Médicales	70	G_004	Autrepart
24	EG_024	Journal Européen des Systèmes Automatisés	71	G_005	Cahiers de la Recherche Architecturale et Urbaine (Les)
25	EG_025	La Revue des Sciences de Gestion	72	G_006	Cahiers des amériques latines
26	EG_026	Logistique & Management	73	G_007	Cahiers d'études africaines
27	EG_027	M@n@gement	74	G_008	Cahiers d'Outre-Mer
28	EG_028	Management & Avenir	75	G_009	Carnets de géographes
29	EG_029	Management et Sciences Sociales	76	G_010	Confins
30	EG_030	Mathématiques et Sciences Humaines	77	G_011	Cybergeog
31	EG_031	Mondes en Développement	78	G_012	EchoGéo
32	EG_032	Négociations	79	G_013	Espace géographique (L')
33	EG_033	Oeconomia	80	G_014	Espace Politique (L')
34	EG_034	Politiques et Management Public	81	G_015	Espace Populations Sociétés
35	EG_035	Question(s) de Management	82	G_016	Espaces et Sociétés
36	EG_036	Recherche et Applications en Marketing	83	G_017	EspacesTemps.net
37	EG_037	Recherches en Sciences de Gestion	84	G_018	Études rurales
38	EG_038	Région et Développement	85	G_019	Field Actions Science Reports
39	EG_039	Réseaux	86	G_020	Flux
40	EG_040	Revue d'Economie du Développement	87	G_021	Géocarrefour
41	EG_041	Revue d'Economie Financière	88	G_022	Géographie et cultures
42	EG_042	Revue d'Economie Industrielle	89	G_023	Géographie, économie, société
43	EG_043	Revue d'Economie Politique	90	G_024	Géomorphologie : relief, processus, environnement
44	EG_044	Revue d'Étude en Agriculture et Environnement / Review of Agricultural, Food and Environmental Studies	91	G_025	Hérodote
45	EG_045	Revue d'études comparatives Est-Ouest	92	G_026	Histoire urbaine
46	EG_046	Revue de Gestion des Ressources Humaines	93	G_027	Hommes & migrations
47	EG_047	Revue de l'Entrepreneuriat	94	G_028	L'Information géographique
48	EG_048	Revue de L'OFCE	95	G_029	M@ppemonde
			96	G_030	Maghreb-Machrek
			97	G_031	Méditerranée

98	G_032	Métropoles	144	H_010	Annales historiques de la Révolution française
99	G_033	Monde Chinois	145	H_011	Annales historiques de l'électricité
100	G_034	Netcom : réseaux, communication et territoires	146	H_012	Année épigraphique (L')
101	G_035	Norois	147	H_013	Antiquités africaines
102	G_036	Outre-Terre : revue française de géopolitique	148	H_014	Archives d'histoire doctrinale et littéraire du Moyen Âge
103	G_037	Photo-Interprétation : European Journal of Applied Remote Sensing	149	H_015	Archives Juives
104	G_038	Physio-Géo	150	H_016	Artefact. Techniques, histoire et sciences humaines
105	G_039	Pôle Sud	151	H_017	Bibliothèque de l'École des chartes
106	G_040	Population	152	H_018	Bulletin de la Société de L'Histoire Du Protestantisme Français (depuis 2016 Revue d'histoire du protestantisme)
107	G_041	Population & Avenir		H_019	Bulletin de la Société d'Histoire de la Guadeloupe
108	G_042	Problèmes d'Amérique Latine		H_020	Bulletin de la Société Nationale des Antiquaires de France
109	G_043	Projets de paysage	153	H_021	Bulletin de l'École française d'Extrême-Orient
110	G_044	Quatenaire	154	H_022	Bulletin de l'Institut Pierre Renouvin
111	G_045	Recherche Transports Sécurité	155	H_023	Bulletin d'études orientales
112	G_046	Revue de Géographie Alpine	156	H_024	Bulletin du Centre d'études médiévales d'Auxerre
113	G_047	Revue d'économie régionale et urbaine	157	H_025	Cahiers alsaciens d'archéologie, d'art et d'Histoire
114	G_048	Revue européenne des migrations internationales	158	H_026	Cahiers de civilisation espagnole contemporaine
115	G_049	Revue géographique de l'Est	161	H_027	Cahiers de Civilisation Médiévale
116	G_050	Revue Internationale de Géomatique	162	H_028	Cahiers de la Méditerranée
117	G_051	Revue internationale des études du développement (Anciennement Revue Tiers-Monde)	163	H_029	Cahiers des amériques latines
118	G_052	Rives méditerranéennes	164	H_030	Cahiers d'Études Hispaniques Médiévales
119	G_053	Science et Changements Planétaires / Sécheresse	165	H_031	Cahiers d'histoire de l'aluminium
120	G_054	Sud-Ouest Européen	166	H_032	Cahiers d'histoire du CNAM : nouvelle série
121	G_055	Territoire en Mouvement	167	H_033	Cahiers d'Histoire et de Philosophie des Sciences
122	E_001	Développement durable et territoires	168	H_034	Cahiers d'histoire. Revue d'histoire critique
123	E_002	ELOHI	169	H_035	Cahiers du Centre Gustave Glotz
124	E_003	Environnement, risques et santé	170	H_036	Cahiers du monde russe
125	E_004	Natures Sciences Sociétés : recherches et débats interdisciplinaires	171	H_037	Chrétiens et sociétés XVIe-XXIe siècles
126	E_005	Revue d'écologie	172	H_038	Chroniques de Port-Royal
127	E_006	Revue juridique de l'environnement	173	H_039	Clio. Femmes, Genre, Histoire
128	E_007	SAPIENS (Surveys and Perspectives Integrating Environment and Society)	174	H_040	Compostelle, Cahiers d'Etudes de Recherche et d'Histoire Compostellanes
129	E_008	Water Alternatives	175	H_041	ComptabilitéS : revue d'histoire des comptabilités
130	E_009	Revue d'ethnoécologie	176	H_042	Comptes-rendus des séances de l'Académie des Inscriptions et Belles-Lettres
131	E_010	Annales des Mines - Responsabilité et environnement	177	H_043	Crime Histoire & Sociétés / Crime, History & Societies
132	E_011	Ecologie & politique		H_044	Critique internationale
133	E_012	Tous urbains		H_045	Diacronie
134	E_013	Climatologie		H_046	Dialogues d'histoire ancienne
135	H_001	1895 : revue d'histoire du cinéma	180	H_047	Diasporas : circulations, migrations, histoire
136	H_002	Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public	181	H_048	Dissidences
137	H_003	Afriques. Débats, méthodes et terrains d'histoire	182	H_049	Dix-huitième siècle
138	H_004	Anabases : Traditions et Réceptions de l'Antiquité	183	H_050	Dix-septième siècle
139	H_005	Annales : Histoire, Sciences Sociales	184	H_051	Écrire l'histoire
140	H_006	Annales de Bretagne et des pays de l'Ouest	185	H_052	Entreprises et histoire
141	H_007	Annales de démographie historique	186	H_053	e-Phaistos : Revue d'histoire des techniques / Journal of History of Technology
142	H_008	Annales de Normandie	187	H_054	Études arméniennes contemporaines
143	H_009	Annales du Midi : Revue archéologique, historique et philologique de la France méridionale	188		

189	H_055	Études balkaniques	237	H_103	Qantara
190	H_056	Études Franciscaines	238	H_104	Recherches internationales
191	H_057	Études normandes	239	H_105	Recherches sur Diderot et sur l'Encyclopédie
192	H_058	Études photographiques	240	H_106	Relations internationales
193	H_059	Europa Moderna - Revue d'histoire et d'iconologie	241	H_107	Révolution française: cahiers de l'Institut d'histoire de la Révolution française
194	H_060	Eurorient	242	H_108	Revue biblique
195	H_061	Extrême-Orient, Extrême-Occident	243	H_109	Revue d'Alsace
196	H_062	Francia. Forschungen zur westeuropäischen Geschichte	244	H_110	Revue d'assyriologie et d'archéologie orientale
197	H_063	Genèses	245	H_111	Revue de l'histoire des religions
198	H_064	Genre & histoire	246	H_112	Revue de l'Institut français d'histoire en Allemagne
199	H_065	Germanica	247	H_113	Revue de Qumrân
200	H_066	Guerres mondiales et conflits contemporains	248	H_114	Revue des études byzantines
201	H_067	Histoire & mesure	249	H_115	Revue des revues : revue internationale d'histoire et de bibliographie
202	H_068	Histoire de la justice	250	H_116	Revue d'études augustinienes et patristiques
203	H_069	Histoire de la recherche contemporaine	251	H_117	Revue d'histoire de la protection sociale
204	H_070	Histoire de l'éducation	252	H_118	Revue d'histoire de la Shoah
205	H_071	Histoire des Sciences médicales	253	H_119	Revue d'histoire de l'Église de France
206	H_072	Histoire et Archives	254	H_120	Revue d'histoire de l'enfance "irrégulière"
207	H_073	Histoire & Sociétés Rurales	255	H_121	Revue d'histoire des chemins de fer
208	H_074	Histoire et Sociétés. Revue Européenne d'Histoire Sociale	256	H_122	Revue d'Histoire des Mathématiques
209	H_075	Histoire Médiévale et Archéologie	257	H_123	Revue d'histoire des sciences
210	H_076	Histoire, économie & société	258	H_124	Revue d'histoire des sciences humaines
211	H_077	Histoire, Epistémologie, Langage (HEL)	259	H_125	Revue d'histoire des textes
212	H_078	Histoire, médecine et santé	260	H_126	Revue d'Histoire diplomatique
213	H_079	Histoire, monde et cultures religieuses	261	H_127	Revue d'Histoire du théâtre
214	H_080	Histoire@Politique	262	H_128	Revue d'histoire du XIXe siècle
215	H_081	Hypothèses. Travaux de l'École doctorale d'histoire de l'université Paris 1 Panthéon-Sorbonne	263	H_129	Revue d'histoire moderne et contemporaine
216	H_082	Cahiers Jaurès	264	H_130	Revue d'Histoire Nordique/Nordic Historical Review
217	H_083	Journal des africanistes	265	H_131	Revue du Nord
218	H_084	Ktéma: Civilisations de l'Orient, de la Grèce et de Rome antiques	266	H_132	Revue française d'histoire des idées politiques
219	H_085	Matériaux pour l'histoire de notre temps	267	H_133	Revue Française d'Histoire du livre
220	H_086	Médiévales : Langue, textes, Histoire	268	H_134	Revue historique
221	H_087	Mélanges de la Casa de Velázquez	269	H_135	Sciences et Techniques en Perspective
222	H_088	Mélanges de l'École française de Rome - Antiquité	270	H_136	Sigila : revue transdisciplinaire franco-portugaise sur le secret
223	H_089	Mélanges de l'École française de Rome - Italie et Méditerranée modernes et contemporaines	271	H_137	Société de l'Histoire de Paris et de l'Ile-de-France. Bulletin
224	H_090	Mélanges de l'École française de Rome - Moyen Âge	272	H_138	Studia Islamica
225	H_091	Mémoires de la Société d'histoire et d'archéologie de Bretagne	273	H_139	Syria. Archéologie, art et histoire
226	H_092	Mil neuf cent : Cahiers Georges Sorel : Revue d'histoire intellectuelle	274	H_140	Témoigner. Entre histoire et mémoire = Getuigen. Tussen Geschiedenis en Herinnering
227	H_093	Monde(s) : histoire, espaces, relations	275	H_141	Le Temps des médias
228	H_094	Mouvement social (Le)	276	H_142	Tracés
229	H_095	Le Moyen Age	277	H_143	Transatlantica
230	H_096	Napoleonica. La Revue	278	H_144	Vingtième Siècle. Revue d'histoire
231	H_097	Nouvelle Revue d'onomastique	279	H_145	XVII-XVIII. Bulletin de la Société d'études anglo-américaines des XVIIe et XVIIIe siècles
232	H_098	Oliviana. Mouvements et dissidences spirituels, XIIIe-XIVe siècles	280	H_146	Yod : revue des études hébraïques et juives
233	H_099	Oriens-Occidens (Cahiers du centre d'Histoire des Sciences et des Philosophies arabes et médiévales)	281	H_147	Zilsel
234	H_100	Outre-mers	282	H_148	Bulletin Monumental
235	H_101	Parlement[s]. Revue d'histoire politique	283	L_001	Amérique : Cahiers du CRICCAL
236	H_102	Provence historique	284	L_002	Année balzacienne (L')
			285	L_003	Année stendhalienne (L')

286	L_004	Arzanà : cahiers de littérature médiévale italienne	336	L_054	Perspectives médiévales - Revue d'épistémologie des langues et littératures du Moyen Âge
287	L_005	Atelier du roman (L')	337	L_055	Po&sie
288	L_006	Austriaca	338	L_056	Poétique
289	L_007	Babel : Littératures plurielles	339	L_057	Présence Africaine
290	L_008	Bien dire et bien apprendre: Revue de médiévistique	340	L_058	Recherches & travaux
291	L_009	Bulletin d'informations proustiennes	341	L_059	Recherches Anglaises et Nord-Américaines
292	L_010	Bulletin de la Société Internationale des Amis de Montaigne	342	L_060	Recherches Germaniques
293	L_011	Bulletin de la Société Internationale des Etudes Yourcenariennes	343	L_061	Réforme Humanisme Renaissance
294	L_012	Bulletin de la Société Paul Claudel	344	L_062	ReS Futurae - Revue d'études sur la science-fiction
295	L_013	Bulletin de l'Association Guillaume Budé	345	L_063	Revue d'études proustiennes
296	L_014	Bulletin des Amis d'André Gide	346	L_064	Revue de littérature comparée
297	L_015	Bulletin hispanique	347	L_065	Revue de philologie, de littérature et d'histoire anciennes
298	L_016	Cahiers d'études germaniques	348	L_066	Revue des Études Anciennes
299	L_017	Cahiers de l'Association Internationale des Etudes Françaises	349	L_067	Revue des Études Grecques
300	L_018	Cahiers de Narratologie	350	L_068	Revue des Etudes Slaves
301	L_019	Cahiers de recherches médiévales et humanistes	351	L_069	Revue des Langues Romanes
302	L_020	Cahiers d'études italiennes	352	L_070	Revue d'histoire littéraire de la France
303	L_021	Cahiers elisabethains	353	L_071	Revue française d'études américaines
304	L_022	Cahiers George Sand	354	L_072	Revue germanique internationale
305	L_023	Cahiers Jean Giraudoux	355	L_073	Revue russe
306	L_024	Cahiers Robinson	356	L_074	Rocamboles (Le). Revue des amis du roman populaire.
307	L_025	Cahiers Valéry Larbaud	357	L_075	Roman 20-50
308	L_026	Cahiers Victoriens and Edouardiens	358	L_076	Romania
309	L_027	Caravelle : Cahiers du Monde hispanique et luso-brésilien	359	L_077	Romantisme
310	L_028	Chroniques italiennes	360	L_078	Rursus : Poétique, réception et réécriture des textes antiques
311	L_029	Elseneur	361	L_079	Seizième Siècle
312	L_030	Etudes anglaises	362	L_080	Siècle 21. Littérature et société
313	L_031	Études britanniques contemporaines	363	L_081	Textes & contextes
314	L_032	Etudes germaniques	364	L_082	TRANS-
315	L_033	Études irlandaises	365	L_083	Viatica
316	L_034	Études littéraires africaines	366	L_084	Vita Latina
317	L_035	Exercices de rhétorique	367	L_085	Théâtre Public
318	L_036	Féeries. Etudes sur le conte merveilleux, XVIIe-XIXe siècle	368	L_086	Ubu: scènes d'Europe - European stages
319	L_037	GAlA. Revue interdisciplinaire sur la Grèce ancienne			
320	L_038	Genesis			
321	L_039	Histoires littéraires			
322	L_040	ILCEA : Revue de l'Institut des langues et cultures d'Europe et d'Amérique			
323	L_041	Infini			
324	L_042	Iris			
325	L_043	Journal of the Short Story in English			
326	L_044	Labyrinthe			
327	L_045	Langues et littératures du monde arabe			
328	L_046	Licorne (La)			
329	L_047	Littérature			
330	L_048	Littératures			
331	L_049	Littératures classiques			
332	L_050	Loxias			
333	L_051	Multitudes			
334	L_052	Nouvelle Revue Française			
335	L_053	Otrante. Arts et littératures fantastiques			

Annexe 2 : Les modes de diffusion des revues françaises de SHS : quelques résultats d'une pré-enquête du Ministère de la Culture

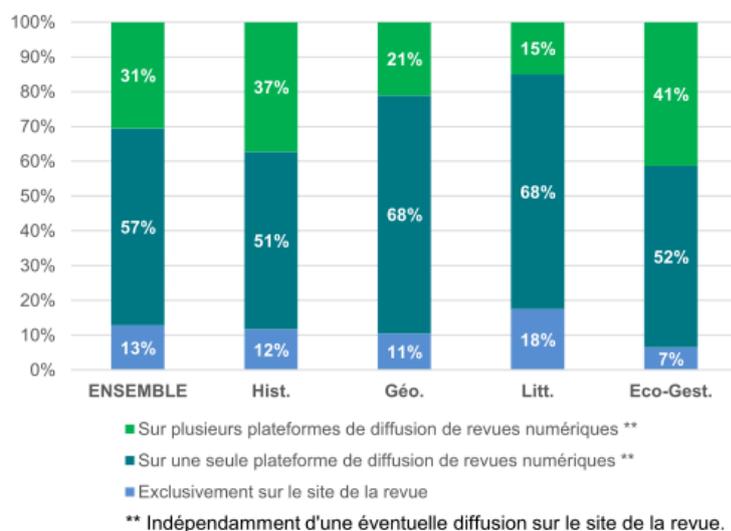
Source: Données extraites d'une étude sur l'économie des revues de SHS commandée par le CSES et diligentée par la Direction Générale des Médias et des Industries Culturelles (DGMIC) et du Service du Livre et de la Lecture (SLL) du Ministère de la culture

Question Q3.1 Votre revue est diffusée...

REPONSES BRUTES	ENSEMBLE	Hist.	Géo.	Litt.	Eco-Gest.
Uniquement sous format papier	13%	15%	3%	23%	6%
Sous format papier et numérique	69%	76%	59%	50%	86%
Désormais uniquement sous format numérique	8%	2%	21%	17%	4%
Depuis l'origine, uniquement sous format numérique	10%	7%	18%	10%	4%
ENSEMBLE	100%	100%	100%	100%	100%
<i>s/t uniquement sous format numérique</i>	19%	9%	38%	27%	8%

Tableau 17 : Mode de diffusion des revues selon le support.

Question Q9.1 : Actuellement, la version numérique de votre revue est diffusée...



- les revues publiées par des associations (20%) et les revues de Littérature (18%) sont les plus nombreuses à se diffuser exclusivement sur leur propre site
- les revues publiées par des éditeurs privés (40%) et les revues d'éco-gestion sont les plus nombreuses à être diffusées sur plusieurs plateformes

Tableau 18 : Répartition des modes de diffusion selon les disciplines

Si Q9.1 = *sur une seule plateforme de diffusion de revues numériques* :

Q9.1.b Sur quelle plateforme ?

Si Q9.1.c = *sur plusieurs plateformes de diffusion de revues numériques* :

Q9.1.c Quelle est aujourd'hui, en termes d'audience, votre principale plateforme de diffusion

	ENSEMBLE	sur une seule plateforme	sur plusieurs plateformes : principale
Cairn	43%	38%	54%
OpenEdition/Revue.org	35%	41%	26%
Persée	9%	8%	10%
Autres	10%	13%	10%
NSP	3%	1%	7%

NB. Total supérieur à 100% (quelques citations de 2 noms de plateformes)

Tableau 19 : principale plateforme de diffusion des revues

REPONSES BRUTES	ENSEMBLE	Hist.	Géo.	Litt.	Eco-Gest.
Uniquement sous format papier	13%	15%	3%	<u>23%</u>	6%
Sous format papier et numérique	69%	76%	59%	50%	86%
Désormais uniquement sous format numérique	8%	2%	<u>21%</u>	<u>17%</u>	4%
Depuis l'origine, uniquement sous format numérique	10%	7%	<u>18%</u>	10%	4%
ENSEMBLE	100%	100%	100%	100%	100%
<i>s/t uniquement sous format numérique</i>	19%	9%	<u>38%</u>	<u>27%</u>	8%

Annexe 3 : Méthode d'extraction OAI-PMH des articles de revues diffusées sur OpenEdition avec R

On commence par installer la librairie « OAIHarvester ». Cette librairie permet d'effectuer les 6 requêtes permises par les entrepôts OAI-PMH, à savoir :

1. *Get Record* : obtenir un enregistrement individuel dans l'entrepôt (par exemple un article avec ses métadonnées : titre, auteur, date de parution, etc.)
2. *Identify* : obtenir des informations générales sur l'entrepôt.
3. *ListIdentifiers* : donne la liste des noms de variables disponibles dans le jeu de données.
4. *ListMetadataFormats* : donne la liste des formats de métadonnées disponibles sur l'entrepôt.
5. *ListRecords* : donne la liste des enregistrements de l'entrepôt.
6. *ListSets* : donne la structure de l'entrepôt, celui-ci étant divisé en « sets ».

Une fois appelée dans le code, on renseigne l'URL de l'entrepôt OAI-PMH sur lequel on va travailler (ici celui d'OpenEdition). Viennent ensuite quelques lignes de codes dont la fonction est d'extraire les bonnes métadonnées. Dans l'étape 6, on spécifie le nom du set concerné (ici celui de la revue Anabases). On extrait alors les métadonnées de ce set qu'on range dans un vecteur « m », que l'on va transformer (« oaih_transform ») dans un format que R peut utiliser (à partir du xml natif) et dont on va extraire (étape 9) les colonnes qui nous intéressent à savoir simplement le titre de l'article, les noms des auteurs et la date de mise en ligne (la date de parution n'est pas disponible dans le jeu de métadonnées au format DublinCore ; pour l'obtenir, il faudrait disposer sur R d'une librairie qui permette d'extraire en OAI-PMH au format METS, ce qui à notre connaissance n'existe pas encore). Enfin, on crée un tableau rassemblant les métadonnées des articles de la revue dans un fichier « csv » (étape 10).

Exemple de code dans R pour l'extraction des articles de la revue Anabases :

1. `library(OAIHarvester)`
2. `baseurl <- "http://oai.openedition.org"`
3. `x <- oaih_identify(baseurl)`
4. `rbind(x, deparse.level = 0)`
5. `oaih_list_sets(baseurl)`
6. `x <- oaih_list_records(baseurl, set = "journals:anabases")`
7. `m <- x[, "metadata"]`
8. `m <- oaih_transform(m[lengths(m) > 0L])`
9. `n <- m[,c(1L,2L,7L)]`
10. `write.csv2(n, "Anabases.csv")`

Annexe 4 : Méthode d'extraction des données de dépôts sur HAL revue par revue

Pour ce faire, nous avons utilisé l'outil de recherche avancée de Hal qui permet d'entrer dans la base de données de Hal à partir d'un nom de revue ou de son identifiant ISSN ou eISSN :

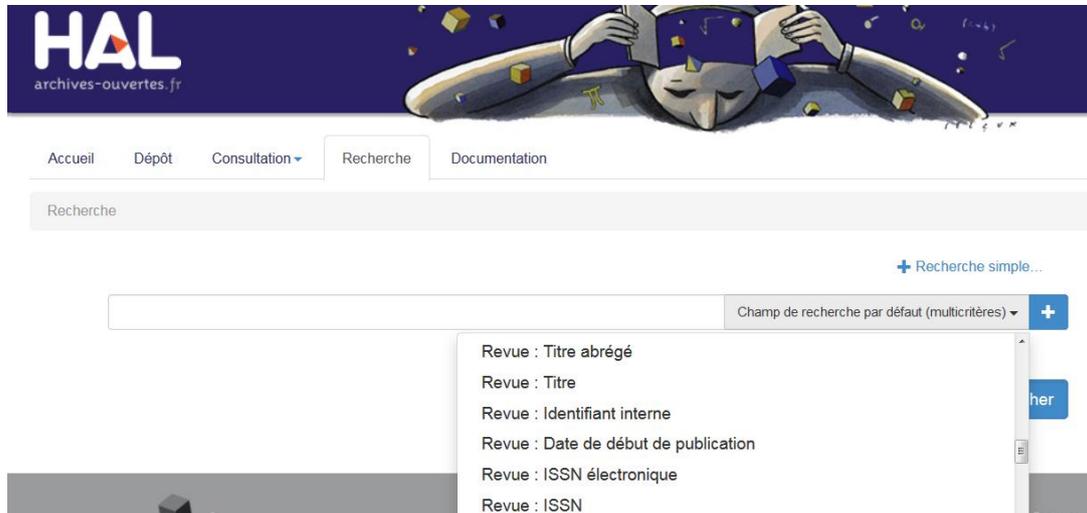


Figure 8 : Capture d'écran de la fonction recherche avancée de Hal par revue

Pour chaque revue de la liste établie par le CSES, nous avons entré son nom ou son ISSN pour obtenir la liste des dépôts effectués. La recherche par « nom de revue » et par « ISSN » donne bien entendu, dans la quasi-totalité des cas, le même résultat. A priori, la recherche par ISSN, qui est un identifiant unique, est d'ailleurs à privilégier car elle permet d'éviter les faux positifs comme les articles de la revue *Population Ecology* qui apparaissent si l'on fait une requête avec le nom de revue *Population*. Toutefois, il arrive que l'ISSN d'une revue ne soit pas renseigné, aussi ne peut-on pas systématiser la recherche par ISSN sans risquer d'omettre des résultats pertinents. Nous avons donc fait le choix de faire une double recherche, par nom de revue et par ISSN, afin de nous assurer de la fiabilité des résultats obtenus.



Figure 9 : Résultats d'une recherche avancée par nom de revue pour la « Revue d'écologie »

Une fois sélectionnés l'intégralité des dépôts (« documents » et « notices ») pour les années de 2010 à 2017, nous préparons l'export des résultats de la recherche à l'aide de l'onglet « outils » situé en

haut à droite (et également en bas à droite) de la colonne des résultats. On sélectionne alors dans le menu déroulant l'option « export avancé » qui permet de choisir précisément les métadonnées que l'on souhaite exporter.

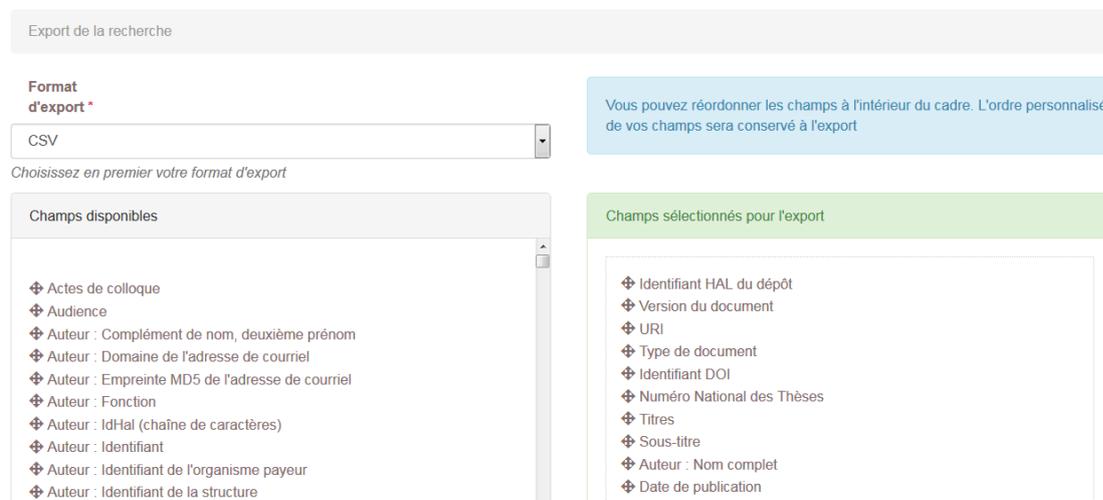


Figure 10 : Capture d'écran de la page de préparation de l'export avancé sur Hal

A droite figurent les champs présélectionnés pour l'export. Y sont renseignés le titre de l'article, son DOI, son auteur, la date de publication, etc. Certaines métadonnées d'intérêt pour notre enquête n'y figurent en revanche pas et il faut les ajouter en faisant glisser les intitulés correspondants du menu « champs disponibles » à gauche au menu « champs sélectionnés pour l'export » à droite. Nous avons ainsi systématiquement ajouté les métadonnées suivantes :

- Date de dépôt sur Hal
- Date de modification du dépôt
- Nom complet du contributeur/déposant
- Type de dépôt (notice ou texte intégral)

La « date de dépôt sur Hal »²³, comparée à la date de parution de l'article, permet d'établir l'intervalle de temps séparant la publication dans la revue académique de sa mise à disposition en libre accès. Ce temps peut ensuite être comparé à la durée de l'embargo pratiquée par la revue afin de vérifier l'impact de la barrière mobile sur la mise à disposition en open access des publications.

Le « nom complet du contributeur/déposant » est indispensable si l'on veut comprendre la diversité des logiques de dépôt et cerner l'importance et les motivations des tiers-déposants (documentalistes, collègues

Chaque champ supplémentaire constitue une colonne de plus dans le tableau qui est exporté au format .csv et qui devient dès lors le matériel de base du travail pour l'analyse des dépôts dans Hal au niveau de granularité de la revue.

²³ Plusieurs formats de date sont disponibles : « date de dépôt », « date de dépôt (ISO 8601) », « date de dépôt : année », « date de dépôt : jour », « date de dépôt : mois ». Par commodité, afin de nous économiser un travail de mise en forme de la date après export, nous avons choisi le format « date de dépôt : année » qui permet de calculer immédiatement, par soustraction de la date de parution de l'article, l'intervalle de temps écoulé depuis sa publication, que nous proposons d'appeler « délai de dépôt en archive ouverte ».

Annexe 5 : Tableaux de synthèse des dépôts dans Hal par discipline (2010-2018)

Annexe 5.1 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Sciences de l'environnement »

Référence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2018	Dépôts dans Hal		Taux de dépôt dans Hal (en %, arrondi au dixième)
						Textes intégraux	Notices	
E_001	Développement durable et territoires	OpenEdition	2010	2017	266	51	83	17,1
E_002	ELOHI	OpenEdition	2012	2016	48	2	4	4,2
E_003	Environnement, risques et santé	Jle	2002	2017	234	8	3	2,9
E_004	Natures Sciences Sociétés : recherches et débats interdisciplinaires	Site web	1993	2017	291	57	106	18,2
E_005	Revue d'écologie	Inist	1931	2017	263	10	42	3,2
E_006	Revue juridique de l'environnement	CAIRN	2010	2018	558	12	58	2,0
E_007	SAPIENS (Surveys and Perspectives Integrating Environment and Society)	Open Edition	2008	2015	60	0	2	0
E_008	Water Alternatives	Site Web	2008	2017	279	3	4	0,0
E_009	Revue d'ethnoécologie	Open Edition	2012	2017	126	11	6	7,6
E_010	Annales des Mines - Responsabilité et environnement	CAIRN	2008	2017	509	10	17	1,7
E_011	Ecologie & politique	CAIRN	2002	2017	204	5	18	2,2
E_012	Tous urbains	CAIRN	2013	2017	143	3	9	1,7
E_013	Climatologie	Inist	2004	2017	51	6	25	11,7

Annexe 5.2 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Géographie »

Référence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2017	Dépôts dans Hal		Taux de dépôt dans Hal (en %, arrondi au dixième)
						Textes intégraux	Notices	
G_001	Annales de géographie	CAIRN	2005	2018	285	24	118	8,4
G_002	Annales de la recherche urbaine	PERSEE	1978	2018	99	15	31	15,1
G_003	ArchéoSciences	OpenEdition	2005	2014	147	14	45	9,5
		CAIRN	2015	2018				
G_004	Autrepart	CAIRN	2001	2018	279	20	56	7,2
G_005	Cahiers de la Recherche Architecturale et Urbaine (Les)	OpenEdition	2009	2018	84	9	8	10,7
G_006	Cahiers des amériques latines	Open Edition	2006	2018	216	28	20	13
G_007	Cahiers d'études africaines	OpenEdition	2000	2015	299	11	47	3,7
		CAIRN	2015	2018				
G_008	Cahiers d'Outre-Mer	CAIRN	2015	2018	204	5	30	2,5
		OpenEdition	2002	2014				
G_009	Carnets de géographes	CAIRN	2013	2018	150	1	43	0,7
G_010	Confins	OpenEdition	2004	2018	498	18	19	3,6
G_011	Cybergeo	OpenEdition	1996	2018	375	68	149	18,1
G_012	EchoGéo	OpenEdition	2007	2018	375	37	129	9,9
G_013	Espace géographique (L')	CAIRN	2001	2018	230	22	94	9,6
G_014	Espace Politique (L')	OpenEdition	2007	2018	243	20	61	8,2
G_015	Espace Populations Sociétés	OpenEdition	2004	2018	218	30	58	13,8
G_016	Espaces et Sociétés	CAIRN	2004	2018	298	23	105	7,7
G_017	EspacesTemps.net	Site Web	2002	2018	175	19	64	10,9
G_018	Études rurales	OpenEdition	2000	2014	171	18	33	10,5
		CAIRN	2015	2018				
G_019	Field Actions Science Reports	OpenEdition	2008	2018	256	4	5	1,6
G_020	Flux	CAIRN	2001	2018	273	29	78	14,6
G_021	Géocarrefour	OpenEdition	2003	2018	198	25	84	12
G_022	Géographie et cultures	OpenEdition	2008	2018	208	38	64	19
G_023	Géographie, économie, société	CAIRN	2004	2018	200	15	80	9,4
G_024	Géomorphologie : relief, processus, environnement	OpenEdition	2005	2018	160	7	96	3,2
G_025	Hérodote	CAIRN	2001	2018	222	11	44	2,9

Référence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2017	Dépôts dans Hal		Taux de dépôt dans Hal (en %, arrondi au dixième)
						Textes intégraux	Notices	
G_026	Histoire urbaine	CAIRN	2000	2018	231	12	47	5,2
G_027	Hommes & migrations	OpenEdition	2009	2014	508	21	101	4,1
		CAIRN	2015	2018				
G_028	L'Information géographique	CAIRN	2006	2018	219	23	58	10,5
G_029	M@ppemonde	Site Web	2004	2018	161	39	36	24,2
G_030	Maghreb-Machrek	CAIRN	2008	2018	180	3	16	1,7
G_031	Méditerranée	OpenEdition	2005	2018	184	6	7	3,3
G_032	Métropoles	OpenEdition	2007	2018	124	20	23	16,1
G_033	Monde Chinois	CAIRN	2012	2018	295	5	5	1,7
G_034	Netcom : réseaux, communication et territoires	OpenEdition	2007	2018	134	21	21	15,7
G_035	Norois	OpenEdition	2004	2014	225	22	79	9,8
		CAIRN	2015	2018				
G_036	Outre-Terre : revue française de géopolitique	CAIRN	2003	2018	793	0	31	0
G_037	Photo-Interprétation	Eska	2008	2018	52	1	20	1,9
G_038	Physio-Géo	OpenEdition	2007	2018	112	11	21	9,8
G_039	Pôle Sud	CAIRN	2004	2018	161	7	24	4,3
G_040	Population	CAIRN	1946	2018	179	3	23	1,7
G_041	Population & Avenir	CAIRN	2005	2018	170	68	15	40
G_042	Problèmes d'Amérique Latine	CAIRN	2008	2018	212	4	24	1,9
G_043	Projets de paysage	Site Web	2008	2018	329	28	34	8,5
G_044	Quatenaire	OpenEdition	2005	2018	250	17	111	6,8
		Springer	2011	2012				
G_045	Recherche Transports Sécurité	Hal	2013	2018	127	87	23	68,5
		OpenEdition	2007	2018				
G_046	Revue de Géographie Alpine	OpenEdition	2007	2018	198	26	54	13,1
G_047	Revue d'économie régionale et urbaine	CAIRN	2001	2018	285	20	105	7
G_048	Revue européenne des migrations internationales	OpenEdition	2002	2017	274	29	42	10,6
		CAIRN	2015	2018				
G_049	Revue géographique de l'Est	OpenEdition	1999	2018	130	10	30	7,7
G_050	Revue Internationale de Géomatique	Revueonline	2001	2018	194	23	73	11,9
G_051	Revue internationale des études du développement	CAIRN	2003	2018	321	21	64	6,5
G_052	Rives méditerranéennes	OpenEdition	1998	2014	349	7	14	4,6
		CAIRN	2015	2018				

Annexe 5.3 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Economie-gestion »

Référence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2017	Dépôts dans Hal		Taux de dépôt dans Hal (en %, arrondi au dixième)
						Textes intégraux	Notices	
EG_001	@GRH	CAIRN	2000	2018	117	7	45	6
EG_002	Actes de la Recherche en Sciences Sociales	CAIRN	1990	2018	257	11	47	4,3
EG_003	Annales d'Economie et de Statistique	JSTOR	2009	2018	246	3	7	1,2
EG_004	Bulletin Français d'Actuariat	Site Web	2003	2018	77	14	2	18,2
EG_005	Cahiers d'Economie Politique	CAIRN	2001	2018	155	4	48	2,6
EG_006	Cahiers scientifiques du transport	Site Web	1979	2018	160	23	19	14,4
EG_007	DigiWorld Economic Journal (anciennement Communications & Stratégies)	IDATE	2015	2018	NR	41	3	NR
EG_008	Comptabilité, Contrôle, Audit	CAIRN	1995	2018	132	6	55	4,5
EG_009	Concurrences	Site Web	2004	2018	160	3	52	1,9
EG_010	Décisions Marketing	EBSCO	2002	2018	237	22	105	9,3
EG_011	Economie et Institutions	OpenEdition	2002	2018	NR	5	14	NR
EG_012	Economie et Prévision	CAIRN	2001	2018	136	10	36	7,4
EG_013	Économie et Statistique	PERSEE	1946	2018	308	38	75	12,3
EG_014	Économie Publique	OpenEdition	1998	2012	22	2	13	9,1
EG_015	Economie Rurale	OpenEdition	2005	2018	297	27	92	9,1
		CAIRN	2015	2014				
EG_016	Entreprendre & Innover	CAIRN	2011	2018	219	5	38	2,3
EG_017	Finance	CAIRN	2005	2018	72	3	18	4,2
EG_018	Finance Contrôle Stratégie	OpenEdition	2012	2018	127	15	60	11,8
EG_019	Gérer et Comprendre	CAIRN	2008	2018	245	14	106	5,7
EG_020	Gestion 2000	CAIRN	2011	2018	282	9	80	3,2
EG_021	Gestion et Management Public	CAIRN	2012	2018	NR	7	28	NR
EG_022	Innovations - Revue d'économie et de management de l'innovation	CAIRN	2001	2018	257	9	37	3,5
EG_023	Journal de Gestion et d'Economie Médicales	CAIRN	2008	2018	161	4	44	2,5
EG_024	Journal Européen des Systèmes Automatisés	Revueonline	2001	2018	NR	36	96	NR
EG_025	La Revue des Sciences de Gestion	CAIRN	2005	2018	458	15	59	3,3

EG_026	Logistique & Management	Taylor & Francis	1994	2018	172	11	55	6,4
EG_027	<u>M@n@gement</u>	CAIRN	2001	2018	195	22	44	11,3
EG_028	Management & Avenir	CAIRN	2004	2018	884	33	300	3,7
EG_029	Management et Sciences Sociales	Harmattan (papier)	2006	2009	0	2		
EG_030	Mathématiques et Sciences Humaines	OpenEdition	1995	2012	68	7	10	10,3
EG_031	Mondes en Développement	CAIRN	2001	2018	312	10	104	3,2
EG_032	Négociations	CAIRN	2004	2018	171	4	19	2,3
EG_033	Oeconomia	OpenEdition	2011	2018	143	4	27	2,8
EG_034	Politiques et Management Public	OpenEdition	2008	2018	198	14	60	7,1
		Revuesonline	2015	2014				
EG_035	Question(s) de Management	CAIRN	2012	2018	173	2	24	1,2
EG_036	Recherche et Applications en Marketing	SAGE	1986	2018	107	14	89	13,1
EG_037	Recherches en Sciences de Gestion	CAIRN	2011	2018	276	8	61	2,9
EG_038	Région et Développement	Site Web	1995	2018	155	7	26	4,5
EG_039	Réseaux	CAIRN	1983	2018	346	42	78	12,2
EG_040	Revue d'Economie du Développement	CAIRN	1983	2018	NR	2	22	NR
EG_041	Revue d'Economie Financière	CAIRN	2011	2018	505	1	82	0,2
		OpenEdition	2006	2014				
EG_042	Revue d'Economie Industrielle	CAIRN	2006	2018	207	12	96	5,8
		OpenEdition	2006	2014				
EG_043	Revue d'Economie Politique	CAIRN	2001	2018	339	20	123	5,9
EG_044	Revue d'Étude en Agriculture et Environnement / Review of Agricultural, Food and Environmental Studies	Ideas REPEC	2008	2015	104	40	48	34,5
		Springer	2016	2018				
EG_045	Revue d'études comparatives Est-Ouest	CAIRN	2009	2018	225	23	48	10,2
EG_046	Revue de Gestion des Ressources Humaines	CAIRN	2008	2018	151	2	29	1,3
EG_047	Revue de l'Entrepreneuriat	CAIRN	2001	2018	145	6	47	4,1
EG_048	Revue de L'OFCE	CAIRN	1990	2018	511	133	44	26
EG_049	Revue de l'Organisation Responsable	CAIRN	2006	2018	93	1	29	1,1
EG_050	Revue de la Régulation	OpenEdition	2007	2018	NR	14	98	NR
EG_051	Revue de Philosophie Economique / Review of Economic Philosophy	CAIRN	2009	2018	102	3	20	2,9
EG_052	Revue Economique	CAIRN	1990	2018	NR	27	203	NR
EG_053	Revue Française d'Economie	CAIRN	2009	2018	227	14	66	6,2
EG_054	Revue Française de Gestion	CAIRN	2002	2018	696	51	258	7,3
EG_055	Revue Française de Gouvernance d'Entreprise	Site Web	2007	2018	109	5		4,6
EG_056	Revue Française d'administration Publique	CAIRN	2002	2018	530	13	73	2,5

EG_057	Revue Française de Socio-économie	CAIRN	2008	2018	228	15	99	6,6
EG_058	Revue Française du Marketing	EBSCO	2001	2018	NR	8	25	NR
EG_059	Revue Interdisciplinaire sur le Management et l'Humanisme (RIMHE)	CAIRN	2012	2018	179	11	52	6,1
EG_060	Revue Internationale de Droit Economique	CAIRN	2001	2018	181	4	39	2,2
EG_061	Revue Internationale de l'Economie Sociale - RECMA	CAIRN	2000	2018	254	5	47	2
EG_062	Revue Internationale de Psychosociologie et de Gestion des Comportements Organisationnels (RIPCO)	CAIRN	2001	2018	245	8	49	3,4
EG_063	Revue internationale des études du développement (anciennement Revue Tiers Monde, avant 2017)	CAIRN	2003	2018	347	NR	NR	NR
EG_064	Sciences Sociales et Santé	CAIRN	2006	2018	198	11	29	5,6
EG_065	Systemes d'Information et Management	CAIRN	2008	2018	145	18	67	12,4
EG_066	Travail et Emploi	CAIRN	2015	2018	195	10	65	5,1
		OpenEdition	2006	2014				

Annexe 5.4 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Histoire »

Réf- érence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2018	Dépôts dans Hal		Taux de présence des articles sur Hal
						Textes intégraux	Notices	
H_001	1895 : revue d'histoire du cinéma	OpenEdition	2000	2018	267	3	3	1,1
H_002	Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public	SHSMESP	2007	2016	181	0	0	0
		Persée	1970	2007	NC			
H_003	Afriques. Débats, méthodes et terrains d'histoire	OpenEdition	2010	2018	74	7	23	9,5
H_004	Anabases : Traditions et Réceptions de l'Antiquité	OpenEdition	2005	2018	311	11	11	3,5
H_005	Annales : Histoire, Sciences Sociales	CAIRN	2001	2018	209	11	13	5,3
H_006	Annales de Bretagne et des pays de l'Ouest	OpenEdition	2009	2014	306	3	20	1
		CAIRN	2015	2018				
H_007	Annales de démographie historique	CAIRN	2001	2018	153	0	16	0
H_008	Annales de Normandie	CAIRN	2010	2018	140	3	17	2,1
H_009	Annales du Midi : Revue archéologique, historique et philologique de la France méridionale	Privat	2014	2018	NR	9	12	NR
		Persée	1989	2013				
H_010	Annales historiques de la Révolution française	OpenEdition			231	8	14	3,5
			1999	2014				
H_011	Annales historiques de l'électricité	CAIRN	2009	2014	45	1	2	2,2
		CAIRN	2003	2008				
H_012	Année épigraphique (L')	CAIRN	2010	2018	248	0	8	0
H_013	Antiquités africaines	Persée	1967	2009	63	1	16	1,6
H_014	Archives d'histoire doctrinale et littéraire du Moyen Âge	CAIRN	2001	2016	63	0	10	0
H_015	Archives Juives	CAIRN	2001	2018	161	1	15	0,6
H_016	Artefact. Techniques, histoire et sciences humaines	OpenEdition	2016	2018	57	3	3	5,3
H_017	Bibliothèque de l'École des chartes	JSTOR	1839	2013	NR	0	1	NR
		Persée	2010	2011				

H_018	Bulletin de la Société de L'Histoire Du Protestantisme Français (depuis 2016 <i>Revue d'histoire du protestantisme</i>)	Droz	2016	2018	208	1	11	0,5
		JSTOR	1903	2015				
H_019	Bulletin de la Société d'Histoire de la Guadeloupe	Erudit	1995	2018	115	1	1	0,9
H_020	Bulletin de la Société Nationale des Antiquaires de France	Persée	1941	2015	289	0	6	0
H_021	Bulletin de l'École française d'Extrême-Orient	Site web	2016	2018	71	11	9	15,5
		Persée	1901	2015				
H_022	Bulletin de l'Institut Pierre Renouvin	CAIRN	2007	2018	166	5	3	3
H_023	Bulletin d'études orientales	OpenEdition	2008	2014	NR	12	8	NR
		CAIRN	2015	2018				
H_024	Bulletin du Centre d'études médiévales d'Auxerre	OpenEdition	2003	2018	231	41	12	17,7
H_025	Cahiers alsaciens d'archéologie, d'art et d'Histoire	Site web	2010	2018	130	0	5	0
H_026	Cahiers de civilisation espagnole contemporaine	OpenEdition	2007	2018	129	1	21	0,8
H_027	Cahiers de Civilisation Médiévale	Persée	1958	2007	NR	3	16	NR
		Site web	2008	2018				
H_028	Cahiers de la Méditerranée	OpenEdition	2001	2018	361	10	8	2,8
H_029	Cahiers des amériques latines	OpenEdition	2006	2018	198	30	7	15,2
H_030	Cahiers d'Études Hispaniques Médiévales	Persée	1976	2015	110	4	12	3,6
		CAIRN	2012	2018				
H_031	Cahiers d'histoire de l'aluminium	CAIRN	2008	2018	76	0	0	0
H_032	Cahiers d'histoire du CNAM : nouvelle série	Site web	2014	2018	59	2	2	3,4
		Site web	1992	1996				
H_033	Cahiers d'Histoire et de Philosophie des Sciences	Site web	1996	2018	0	0	0	0
H_034	Cahiers d'histoire. Revue d'histoire critique	OpenEdition	2001	2018	239	10	9	4,2
H_035	Cahiers du Centre Gustave Glotz	Site web	1991	2016	129	7	NR	5,4
H_040	Compostelle, Cahiers d'Etudes de Recherche et d'Histoire Compostellanes	Site web			25	0	0	0
H_041	ComptabilitÉS : revue d'histoire des comptabilités	OpenEdition			83	2	15	2,4
H_042	Comptes-rendus des séances de l'Académie des Inscriptions et Belles-Lettres	Site web	2008	2015	418	14	9	3,3
		Persée	1857	2011				
H_043	Crime Histoire & Sociétés / Crime, History & Societies	OpenEdition	1997	2018	124	1	14	0,8
H_044	Critique internationale	CAIRN	1998	2018	368	26	19	7,1

H_045	Diacronie	OpenEdition	2011	2016	NR	0	2	NR
H_046	Dialogues d'histoire ancienne	CAIRN	2007	2018	397	4	16	1
H_047	Diasporas : circulations, migrations, histoire	OpenEdition	2012	2018	201	5	4	2,5
H_048	Dissidences	Hypotheses			0	0	15	#DIV/0!
H_049	Dix-huitième siècle	CAIRN	2006	2018	334	8	9	2,4
H_050	Dix-septième siècle	CAIRN	2001	2018	381	3	12	0,8
H_051	Écrire l'histoire	OpenEdition	2009	2018	200	6	22	3
H_052	Entreprises et histoire	CAIRN	1992	2018	359	17	17	4,7
H_053	e-Phaistos : Revue d'histoire des techniques / Journal of History of Technology	OpenEdition	2012	2018	66	1	0	1,5
H_054	Études arméniennes contemporaines	OpenEdition	2012	2018	65	1	2	1,5
H_055	Études balkaniques	CAIRN	1994	2018	59	1	0	1,7
H_056	Etudes Franciscaines	Site web	2008	2018	109	3	9	2,8
H_057	Études normandes	Site web	1951	2018	197	3	7	1,5
H_058	Études photographiques	OpenEdition	1996	2018	128	10	13	7,8
H_059	Europa Moderna - Revue d'histoire et d'iconologie	Site web	2010	2018	NR	1	8	NR
H_060	Eurorient	Site web	2008	2014	160	2	2	1,3
H_061	Extrême-Orient, Extrême-Occident	CAIRN	2015	2018	108	6	27	5,6
		OpenEdition	2008	2014				
H_062	Francia. Forschungen zur westeuropäischen Geschichte	Site web	1973	2018	0	1	6	#DIV/0!
H_063	Genèses	CAIRN	2000	2018	324	18	32	5,6
H_064	Genre & histoire	OpenEdition	2007	2018	104	3	9	2,9
		OpenEdition	1988	2014				
H_065	Germanica	CAIRN	2015	2018	204	1	14	0,5
		CAIRN	2015	2018				
H_066	Guerres mondiales et conflits contemporains	CAIRN	2001	2018	294	3	19	1
H_067	Histoire & mesure	CAIRN	2015	2018	108	3	12	2,8
		OpenEdition	2001	2014				
H_068	Histoire de la justice	CAIRN	2005	2018	135	1	6	0,7
H_069	Histoire de la recherche contemporaine	OpenEdition	2012	2016	NR	3	17	NR
H_070	Histoire de l'éducation	CAIRN	2015	2018	117	1	17	0,9
		OpenEdition	2000	2014				

H_071	Histoire des Sciences médicales	Site web	1967	2018	385	4	31	1
H_072	Histoire et Archives	Site web	1997	2002	0	0	0	#DIV/0!
H_073	Histoire & Sociétés Rurales	CAIRN	2001	2018	97	0	2	0
H_074	Histoire et Sociétés. Revue Européenne d'Histoire Sociale	Site web	2002	2009	0	0	0	#DIV/0!
H_075	Histoire Médiévale et Archéologie	Site web	1988	2018	NR	1	1	NR
H_076	Histoire, économie & société	CAIRN	2004	2018	262	12	9	4,6
H_077	Histoire, Epistémologie, Langage (HEL)	Persée	1979	2015	0	9	17	#DIV/0!
		Site web	2015	2018				
H_078	Histoire, médecine et santé	OpenEdition	2012	2018	87	5	15	5,7
H_079	Histoire, monde et cultures religieuses	CAIRN	2007	2018	168	3	26	1,8
H_080	Histoire@Politique	CAIRN	2007	2018	247	25	12	10,1
H_081	Hypothèses. Travaux de l'École doctorale d'histoire de l'université Paris 1 Panthéon-Sorbonne	Site web	1997	2018	241	3	2	1,2
H_082	Cahiers Jaurès	CAIRN	2001	2018	242	3	18	1,2
H_083	Journal des africanistes	OpenEdition	2004	2018	170	7	16	4,1
H_084	Ktéma: Civilisations de l'Orient, de la Grèce et de Rome antiques	Site web	1976	2018	190	28	6	14,7
H_085	Matériaux pour l'histoire de notre temps	CAIRN	2005	2018	194	3	15	1,5
		OpenEdition	2003	2014				
H_086	Médiévales : Langue, textes, Histoire	CAIRN	2015	2018	192	8	2	4,2
		OpenEdition	2003	2018				
H_087	Mélanges de la Casa de Velázquez	OpenEdition	2003	2018	213	13	17	6,1
H_088	Mélanges de l'École française de Rome - Antiquité	OpenEdition	2010	2018	231	11	9	4,8
H_089	Mélanges de l'École française de Rome - Italie et Méditerranée modernes et contemporaines	OpenEdition			285	3	3	1,1
H_090	Mélanges de l'École française de Rome - Moyen Âge	OpenEdition	2010	2018	302	3	19	1
H_091	Mémoires de la Société d'histoire et d'archéologie de Bretagne	Site web	1920	2018	NR	1	9	NR
H_092	Mil neuf cent : Cahiers Georges Sorel : Revue d'histoire intellectuelle	Site web	1983	2018	66	0	3	0
H_093	Monde(s) : histoire, espaces, relations	CAIRN	2012	2018	155	6	9	3,9
H_094	Mouvement social (Le)	CAIRN	1990	2018	284	10	71	3,5
H_095	Le Moyen Age	CAIRN	2001	2018	154	2	8	1,3
H_096	Napoleonica. La Revue	CAIRN	2008	2018	130	0	0	0
H_097	Nouvelle Revue d'onomastique	Persée	1983	2015	88	2	6	2,3

H_098	Oliviana. Mouvements et dissidences spirituels, XIIIe-XIVe siècles	OpenEdition	2003	2016	26	0	3	0
H_099	Oriens-Occidens (Cahiers du centre d'Histoire des Sciences et des Philosophies arabes et médiévales)	Site web			NR	2	0	NR
H_100	Outre-mers	Site web	1999	2018	310	6	11	1,9
H_101	Parlement[s]. Revue d'histoire politique	CAIRN	2004	2018	298	2	24	0,7
H_102	Provence historique	Site web	1950	2014	NR	6	10	NR
H_103	Qantara	Site web			0	3	13	#DIV/0!
H_104	Recherches internationales	Site web			NR	5	15	NR
H_105	Recherches sur Diderot et sur l'Encyclopédie	CAIRN	2015	2018	142	4	17	2,8
		OpenEdition	1998	2014				
H_106	Relations internationales	CAIRN	2005	2018	293	6	17	2
H_107	Révolution française: cahiers de l'Institut d'histoire de la Révolution française	OpenEdition	2012	2018	140	7	126	5
H_108	Revue biblique	Site web	2010	2018	249	0	4	0
H_109	Revue d'Alsace	OpenEdition	2006	2018	197	0	2	0
H_110	Revue d'assyriologie et d'archéologie orientale	CAIRN	2001	2018	107	0	2	0
H_111	Revue de l'histoire des religions	CAIRN	2015	2018	217	3	18	1,4
		OpenEdition	2005	2014				
H_112	Revue de l'Institut français d'histoire en Allemagne	OpenEdition	2009	2014	28	1	8	3,6
H_113	Revue de Qumrân	Site web	1958	2016	NR	0	4	NR
H_114	Revue des études byzantines	Site web	2014	2018	97	1	7	1
		Persée	1897	2013				
H_115	Revue des revues : revue internationale d'histoire et de bibliographie	CAIRN	2013	2018	152	0	0	0
H_116	Revue d'études augustinienne et patristiques	Brepols	2010	2018	116	2	4	1,7
H_117	Revue d'histoire de la protection sociale	CAIRN	2008	2018	98	2	7	2
H_118	Revue d'histoire de la Shoah	CAIRN	2005	2018	398	0	13	0
H_119	Revue d'histoire de l'Église de France	Brepols	2010	2018	139	4	15	2,9
H_120	Revue d'histoire de l'enfance "irrégulière"	OpenEdition	1998	2018	91	2	12	2,2
H_121	Revue d'histoire des chemins de fer	OpenEdition	2002	2018	NR	4	0	NR
H_122	Revue d'Histoire des Mathématiques	Site web	1995	2018	57	1	15	1,8
H_123	Revue d'histoire des sciences	CAIRN	2006	2018	123	5	7	4,1

H_124	Revue d'histoire des sciences humaines	CAIRN	1999	2011	NR	16	9	NR
		OpenEdition	2016	2018				
H_125	Revue d'histoire des textes	Brepols	2010	2018	112	1	8	0,9
H_126	Revue d'Histoire diplomatique	Site web	2012	2018	NR	1	5	NR
H_127	Revue d'Histoire du théâtre	Site web	2011	2018	311	9	20	2,9
H_128	Revue d'histoire du XIXe siècle	CAIRN	2015	2018	NR	8	16	NR
		OpenEdition	1985	2014				
H_129	Revue d'histoire moderne et contemporaine	CAIRN	1977	2018	276	9	73	3,3
H_130	Revue d'Histoire Nordique/Nordic Historical Review	Site web			NR	7	8	NR
H_131	Revue du Nord	CAIRN	2001	2018	337	6	43	1,8
H_132	Revue française d'histoire des idées politiques	CAIRN	2001	2018	181	3	22	1,7
H_133	Revue Française d'Histoire du livre	Site web	1931	2018	88	4	5	4,5
H_134	Revue historique	CAIRN	1991	2018	223	5	54	2,2
H_135	Sciences et Techniques en Perspective	Blanchard	1997	2018	NR	0	1	NR
H_136	Sigila : revue transdisciplinaire franco-portugaise sur le secret	Site web	1998	2018	318	0	0	0
H_137	Societe de l'Histoire de Paris et de l'Ile-de- France. Bulletin	Gallica	1874	2007	NR	3	2	NR
		Site web	1874	2018				
H_138	Studia Islamica	JSTOR	1953	2015	51	3	2	5,9
H_139	Syria. Archéologie, art et histoire	OpenEdition	2006	2018	156	13	18	8,3
H_140	Témoigner. Entre histoire et mémoire = Getuigen. Tussen Geschiedenis en Herinnering	OpenEdition	2013	2018	375	0	10	0
H_141	Le Temps des médias	CAIRN	2003	2018	289	10	59	3,5
H_142	Tracés	OpenEdition	2002	2018	325	19	62	5,8
H_143	Transatlantica	OpenEdition	2001	2018	357	10	53	2,8
H_144	Vingtième Siècle. Revue d'histoire	CAIRN	1990	2018	411	14	6	3,4
H_145	XVII-XVIII. Bulletin de la Société d'études anglo-américaines des XVIIe et XVIIIe siècles	OpenEdition	2012	2018	148	0	1	0
H_146	Yod : revue des études hébraïques et juives	OpenEdition	2009	2018	75	1	0	1,3
H_147	Zilsel	CAIRN	2017	2018	72	1	11	1,4
H_148	Bulletin Monumental	Site web	2009	2018	130	8	89	6,2

Annexe 5.5 Tableau de synthèse des dépôts dans Hal pour les revues de la discipline « Littérature »

Réf- érence	Nom de la revue	Diffusion en ligne	Début de la collection en ligne	Fin de la collection en ligne	Nombre d'articles de recherche publiés de 2010 à 2017	Dépôts dans Hal		Taux de présence des articles sur Hal
						Textes intégraux	Notices	
L_001	Amérique : Cahiers du CRICCAL	OpenEdition	2011	2018	169			0,6
L_002	Année balzacienne (L')	CAIRN	2000	2017	173	0	6	0
L_003	Année stendhalienne (L')	Honoré Champion			168	3	9	1,8
L_004	Arzanà : cahiers de littérature médiévale italienne	OpenEdition	2010	2017	81	6		7,4
L_005	Atelier du roman (L')	Site web	1993	2017	NR	0	9	NR
L_006	Austriaca	OpenEdition			NR	1	11	NR
L_007	Babel : Littératures plurielles	OpenEdition	1996	2017	246	7		2,8
L_008	Bien dire et bien apprendre: Revue de médiévistique	Site web			89	2	7	2,9
L_009	Bulletin d'informations proustiennes	Presses ENS			90	2	3	2,2
L_010	Bulletin de la Société Internationale des Amis de Montaigne	Site web			125	1	19	0,8
L_011	Bulletin de la Société Internationale des Etudes Yourcenariennes	Site web			56	0		0
L_012	Bulletin de la Société Paul Claudel	Site web			51	1	0	2
L_013	Bulletin de l'Association Guillaume Budé	Site web			92	3	10	3,3
L_014	Bulletin des Amis d'André Gide	Site web	1985	2017	78	2	4	2,6
L_015	Bulletin hispanique	OpenEdition	2007	2018	291	1	11	0,3
L_016	Cahiers d'études germaniques	OpenEdition	2014	2017	123	14	27	11,4
L_017	Cahiers de l'Association Internationale des Etudes Françaises	Site web			154	1	6	0,6
L_018	Cahiers de Narratologie	OpenEdition	2001	2018	165	11	17	6,7
L_019	Cahiers de recherches médiévales et humanistes	OpenEdition	1996	2016	309	11	58	3,6
L_020	Cahiers d'études italiennes	OpenEdition	2004	2017	200	3	28	1,5
L_021	Cahiers elisabethains	SAGE	1984	2017	130	0	10	0
L_022	Cahiers George Sand	Site web			80	2	0	2,5
L_023	Cahiers Jean Giraudoux	LCDPU			73	1	0	1,4

L_024	Cahiers Robinson	Site web	1997	2017	232	4	22	1,7
L_025	Cahiers Valéry Larbaud	Classiques Garnier			43	0	3	0
L_026	Cahiers Victoriens and Edouardiens	OpenEdition	2010	2017	200	7		3,5
L_027	Caravelle : Cahiers du Monde hispanique et luso-brésilien	OpenEdition	2011	2017	141	11		7,8
L_028	Chroniques italiennes	Site web	1984	2016	177	14		7,9
L_029	Elseneur	LCDPU	1995	2017	87	4	14	4,6
L_030	Etudes anglaises	CAIRN	2001	2017	212	3	54	1,4
L_031	Études britanniques contemporaines	OpenEdition	2015	2017	214	7	47	3,3
L_032	Etudes germaniques	CAIRN	2006	2017	321	0	15	0
L_033	Études irlandaises	OpenEdition	2009	2017	185	4	13	2,2
L_034	Études littéraires africaines	Erudit	1996	2017	156	0	35	0
L_035	Exercices de rhétorique	OpenEdition	2013	2017	84	1	23	1,2
L_036	Féeries. Etudes sur le conte merveilleux, XVIIe-XIXe siècle	OpenEdition	2004	2017	85	2	10	2,4
L_037	GAIA. Revue interdisciplinaire sur la Grèce ancienne	Persee	1997	2016	124	4	12	3,2
L_038	Genesis	OpenEdition	2010	2017	184	4	17	2,2
L_039	Histoires littéraires	Site web	2000	2017	217	0	7	0
L_040	ILCEA : Revue de l'Institut des langues et cultures d'Europe et d'Amérique	OpenEdition	2002	2017	203	25		12,3
L_041	Infini	Gallimard			351	0	1	0
L_042	Iris	LCDPU	1989	2017	NR	0	1	NR
L_043	Journal of the Short Story in English	OpenEdition	1997	2017	181	3	18	1,7
L_044	Labyrinthe	OpenEdition	1998	2017	84	1	3	1,2
L_045	Langues et littératures du monde arabe	Site web	2000	2017	22	0	4	0
L_046	Licorne (La)	PUR	2001	2017	587	3	34	0,5
		Site web	1976	2000				
L_047	Littérature	CAIRN	2006	2017	259	12	37	4,6
L_048	Littératures	OpenEdition	2011	2016	187	4	12	2,1
L_049	Littératures classiques	CAIRN	2004	2017	383	13	1	3,4
L_050	Loxias	Site web	2003	2017	404	6	21	1,5
L_051	Multitudes	Site web			623	17	5	2,7

		CAIRN	2010	2017				
		Scopalto	2011	2017				
L_052	Nouvelle Revue Francaise	Site web	1908	2017	559	0	8	0
L_053	Otrante. Arts et littératures fantastiques	Site web			NR	0	22	NR
L_054	Perspectives médiévales - Revue d'épistémologie des langues et littératures du Moyen Âge	OpenEdition	2009	2017	54	5	8	9,3
L_055	Po&sie	CAIRN	2005	2017	518	0	3	0
L_056	Poétique	CAIRN	2003	2017	160	1	14	0,6
L_057	Présence Africaine	CAIRN	1947	2016	233	0	3	0
L_058	Recherches & travaux	OpenEdition	2005	2017	170	2	9	1,2
L_059	Recherches Anglaises et Nord-Américaines	LCDPU	2000	2017	111	3	18	2,7
L_060	Recherches Germaniques	OpenEdition	2011	2017	96	2	11	2,1
L_061	Réforme Humanisme Renaissance	PERSEE	1975	2014	142	0	12	0
		CAIRN	2015	2017				
L_062	ReS Futurae - Revue d'études sur la science-fiction	OpenEdition	2012	2017	86	4	19	4,7
L_063	Revue d'études proustiennes	Classiques Garnier	2015	2017	141	0	2	0
L_064	Revue de littérature comparée	CAIRN	2001	2017	220	6	14	2,7
L_065	Revue de philologie, de littérature et d'histoire anciennes	CAIRN	2001	2016	121	0	10	0
L_066	Revue des Études Anciennes	Pub editions			138	NR	70	NR
		Site web						
L_067	Revue des Études Grecques	Persee	1888	2016	98	2	8	2
L_068	Revue des Etudes Slaves	Persee	1921	2012	277	10	24	3,6
L_069	Revue des Langues Romanes	OpenEdition	2015	2017	63	2	13	3,2
L_070	Revue d'histoire littéraire de la France	CAIRN	1982	2016	314	6	49	1,9
L_071	Revue française d'études américaines	CAIRN	2001	2017	267	7		2,6
L_072	Revue germanique internationale	OpenEdition	1994	2017	189	1	9	0,5
L_073	Revue russe	Persee	1991	2013	75	6	17	8
L_074	Rocamboles (Le). Revue des amis du roman populaire.	Site web	1997	2017	301	1	2	0,3
L_075	Roman 20-50	CAIRN	2005	2017	224	5	16	2,2
L_076	Romania	Droz	2013	2017	122	1	10	0,8

		Persee	1872	2011	32			0
L_077	Romantisme	CAIRN	2004	2017	307	15	49	4,9
L_078	Rursus : Poétique, réception et réécriture des textes antiques	OpenEdition	2006	2017	39	4	5	10,3
L_079	Seizième Siècle	Persee	2005	2014	85	1	29	1,2
L_080	Siècle 21. Littérature et société	Site web	2002	2017	630	0	1	0
L_081	Textes & contextes	Site web	2008	2014	156	38	27	24,4
L_082	TRANS-	OpenEdition	2005	2017	192	1	11	0,5
L_083	Viatica	Site web	2014	2017	51	4	9	7,8
L_084	Vita Latina	Persee	1989	2014	67	3	15	4,5
L_085	Théâtre Public	Editions théâtrales	1974	2017	0	8	45	#DIV/0!
L_086	Ubu: scènes d'Europe - European stages	Site web	2016	2017	162			

Annexe 6 : script Python pour l'identification de la présence des articles sur Academia et ResearchGate

Ces scripts sont le résultat du travail de Jacques Hilbey et Valentin Barrère, respectivement élève et intervenant au sein du Master MÉDAS (Méga-Données et Analyse Sociale) du CNAM.

Ci-dessous est détaillé le script réalisé pour ResearchGate, avec des fonctionnalités de gestion des erreurs. Celui utilisé pour Academia reprend exactement la même structure.

1. Préambule avec librairies à importer :

```
import urllib.request
from urllib.request import Request, urlopen
import re
from bs4 import BeautifulSoup
import pandas as pd
import subprocess
import requests
import unicodedata
import time
import numpy as np
import os
```

2. Chargement du fichier titre auteur date pour la géographie :

```
dfArt=pd.read_csv('BDD_titre-auteur-date_geographie.csv', sep=';', header=0,
encoding="ansi")
```

3. Puis nettoyage des balises html dans les titres

```
dfArt['TITRE_ARTICLE'] = dfArt['TITRE_ARTICLE'].str.replace('<[^>*>', '')
```

4. Définition d'une fonction DfArt par revue de géographie :

```
dfArtGeo1=dfArt.loc[dfArt[""REVUE""]==""Annales de géographie"",:]
dfArtGeo3=dfArt.loc[dfArt[""REVUE""]==""Archéosciences"",:]
dfArtGeo4=dfArt.loc[dfArt[""REVUE""]==""Autrepart"",:]
dfArtGeo5=dfArt.loc[dfArt[""REVUE""]==""Cahiers de la recherche architecturale et
urbaine"",:]
dfArtGeo6=dfArt.loc[dfArt[""REVUE""]==""Cahiers des amériques latines"",:]
dfArtGeo8=dfArt.loc[dfArt[""REVUE""]==""Cahiers d'Outre-Mer"",:]
dfArtGeo9=dfArt.loc[dfArt[""REVUE""]==""Carnets de géographes"",:]
dfArtGeo10=dfArt.loc[dfArt[""REVUE""]==""Confins"",:]
dfArtGeo11=dfArt.loc[dfArt[""REVUE""]==""Cybergeographie"",:]
dfArtGeo12=dfArt.loc[dfArt[""REVUE""]==""ECHOgéographie"",:]
dfArtGeo13=dfArt.loc[dfArt[""REVUE""]==""Espace géographique"",:]
dfArtGeo14=dfArt.loc[dfArt[""REVUE""]==""Espace politique"",:]
dfArtGeo15=dfArt.loc[dfArt[""REVUE""]==""Espace populations sociétés"",:]
dfArtGeo16=dfArt.loc[dfArt[""REVUE""]==""Espaces et sociétés"",:]
dfArtGeo17=dfArt.loc[dfArt[""REVUE""]==""Espacetemps.net"",:]
dfArtGeo18=dfArt.loc[dfArt[""REVUE""]==""Etudes rurales"",:]
dfArtGeo19=dfArt.loc[dfArt[""REVUE""]==""Field Actions Science Reports"",:]
dfArtGeo20=dfArt.loc[dfArt[""REVUE""]==""Flux"",:]
dfArtGeo21=dfArt.loc[dfArt[""REVUE""]==""Géocarrefour"",:]
dfArtGeo22=dfArt.loc[dfArt[""REVUE""]==""Géographie et cultures"",:]
dfArtGeo23=dfArt.loc[dfArt[""REVUE""]==""Géographie, économie, société"",:]
```

```

dfArtGeo24=dfArt.loc[dfArt[""REVUE""]=="Géomorphologie : relief, processus,
environnement",:]
dfArtGeo25=dfArt.loc[dfArt[""REVUE""]=="Hérodote",:]
dfArtGeo26=dfArt.loc[dfArt[""REVUE""]=="Histoire urbaine",:]
dfArtGeo27=dfArt.loc[dfArt[""REVUE""]=="Hommes & migrations",:]
dfArtGeo28=dfArt.loc[dfArt[""REVUE""]=="L'Information géographique",:]
dfArtGeo29=dfArt.loc[dfArt[""REVUE""]=="M@ppemonde",:]
dfArtGeo30=dfArt.loc[dfArt[""REVUE""]=="Maghreb-Machrek",:]
dfArtGeo31=dfArt.loc[dfArt[""REVUE""]=="Méditerranée",:]
dfArtGeo32=dfArt.loc[dfArt[""REVUE""]=="Métropoles",:]
dfArtGeo33=dfArt.loc[dfArt[""REVUE""]=="Monde Chinois",:]
dfArtGeo34=dfArt.loc[dfArt[""REVUE""]=="Netcom",:]
dfArtGeo35=dfArt.loc[dfArt[""REVUE""]=="Norois",:]
dfArtGeo36=dfArt.loc[dfArt[""REVUE""]=="Outre-Terre",:]
dfArtGeo37=dfArt.loc[dfArt[""REVUE""]=="Photo-Interprétation",:]
dfArtGeo38=dfArt.loc[dfArt[""REVUE""]=="Physio-Géo",:]
dfArtGeo39=dfArt.loc[dfArt[""REVUE""]=="Pôle Sud",:]
dfArtGeo40=dfArt.loc[dfArt[""REVUE""]=="Population",:]
dfArtGeo41=dfArt.loc[dfArt[""REVUE""]=="Population & Avenir",:]
dfArtGeo42=dfArt.loc[dfArt[""REVUE""]=="Problèmes d'Amérique Latine",:]
dfArtGeo43=dfArt.loc[dfArt[""REVUE""]=="Monde Chinois",:]
dfArtGeo44=dfArt.loc[dfArt[""REVUE""]=="Quaternaire",:]
dfArtGeo46=dfArt.loc[dfArt[""REVUE""]=="Revue de géographie alpine",:]
dfArtGeo47=dfArt.loc[dfArt[""REVUE""]=="Revue d'économie régionale et
urbaine",:]
dfArtGeo48=dfArt.loc[dfArt[""REVUE""]=="Revue européenne des migrations
internationales",:]
dfArtGeo49=dfArt.loc[dfArt[""REVUE""]=="Revue géographique de l'Est",:]
dfArtGeo52=dfArt.loc[dfArt[""REVUE""]=="Rives méditerranéennes",:]

```

5. Visualisation d'un aperçu du tableau correspondant à une revue : ici Autrepart, soit DfArtGeo4

```
dfArtGeo4.head()
```

	Code revue	REVUE	TITRE_ARTICLE	Auteur	date de mise en ligne
614	G_004	Autrepart	État développeur, État fragile : comment l'île...	Emmanuel Grégoire	07/12/2017
615	G_004	Autrepart	Vieillesse démographique en Chine : perspe...	Isabelle Attané	07/12/2017
616	G_004	Autrepart	Les mobilités quotidiennes, un révélateur des ...	Florence Boyer	07/12/2017
617	G_004	Autrepart	Emploi domestique et travail identitaire chez ...	Maud Laëthier	07/12/2017
618	G_004	Autrepart	Le registre péruvien de savoirs collectifs des...	Adriana Munoz Sanchez	07/12/2017

6. Définition d'une fonction qui interroge ResearchGate

```

def urlGate(article):
    titre = re.findall(r'(\w+)', article, re.UNICODE)
    titre = ' '.join(titre) # on récupère le titre initial sans ponctuation
    titre = re.sub(' ', '%252B', titre)
    titre = re.sub('à', '%25C3%25A0', titre)
    titre = re.sub('â', '%25C3%25A2', titre)
    titre = re.sub('é', '%25C3%25A9', titre)
    titre = re.sub('É', '%25C3%2589', titre)

```

```

titre = re.sub('È', '%25C3%2588', titre)
titre = re.sub('É', '%25C3%258a', titre)
titre = re.sub('è', '%25C3%25A8', titre)
titre = re.sub('ê', '%25C3%25AA', titre)
titre = re.sub('ë', '%25C3%25AB', titre)
titre = re.sub('î', '%25C3%25AE', titre)
titre = re.sub('ï', '%25C3%25AF', titre)
titre = re.sub('ñ', '%25C3%25B1', titre)
titre = re.sub('ò', '%25C3%25B2', titre)
titre = re.sub('ó', '%25C3%25B3', titre)
titre = re.sub('ô', '%25C3%25B4', titre)
titre = re.sub('õ', '%25C3%25B5', titre)
titre = re.sub('ö', '%25C3%25B6', titre)
titre = re.sub('ù', '%25C3%25B9', titre)
titre = re.sub('ü', '%25C3%25BC', titre)
titre = re.sub('ç', '%25C3%25A7', titre)
titre = re.sub('Ã', '%25C3%2582', titre)
titre = re.sub('æ', '%25C3%25A6', titre)
titre = re.sub('À', '%25C3%2580', titre)
titre = re.sub('Á', '%25C3%2581', titre)
req='https://www.researchgate.net/search/publications?q=%252522'+titre+'%252522'
return(req)

```

```

def reqrepResGate(article):
    res=0
    req = urlGate(article)
    webpage = urllib.request.urlopen(req)
    soup = BeautifulSoup(webpage, "lxml")
    titre2 = unicodedata.normalize('NFD', article).encode('ascii', 'ignore')
    titre2=titre2.decode('ascii')
    titre2='_'.join(re.findall(r'(\w+)', titre2, re.UNICODE))
    for a in soup.find_all('a', href=True):
        if titre2 in a['href']:
            res=1
            break
        else:
            pass
    time.sleep(3)
    return(res)

```

7. Définition d'une autre fonction au même rôle avec boucle

```

def create_csv_avec_loop(dfArtGeo, name_csv):
    """
    Meme chose mais avec une boucle
    """

    list_RG_responses = []
    bool_stop = False
    nb_articles = len(dfArtGeo.TITRE_ARTICLE.values)
    for idx_art_name, art_name in enumerate(dfArtGeo.TITRE_ARTICLE.values):
        print('Article "%s..." : %d/%d' % (art_name[:15], idx_art_name, nb_articles))
        if not bool_stop:
            try:
                list_RG_responses.append((art_name, reqrepResGate(art_name)))
            except:
                csv_name_when_error = name_csv[:-4] + 'erreur.csv'
                print('Probleme, on ecrit un csv %s' % csv_name_when_error)
                pd.DataFrame(list_RG_responses, columns = ['TITRE_ARTICLE',
'Researchgate']).to_csv(csv_name_when_error, encoding = 'utf-8')
                bool_stop = True

```

```

if not bool_stop:
    df_art = pd.DataFrame(list_RG_responses, columns = ['TITRE_ARTICLE',
'Researchgate'])
    dfArtGeo['Researchgate'] = df_art['Researchgate']

    dfArtGeo.to_csv(dfArtGeo, encoding = 'utf-8')

```

8. On teste le résultat sur une revue:

```

# test pour le fichier
dftest = dfArtGeo4.loc[165:170].copy()
dftest = dftest.assign(Researchgate = dftest.TITRE_ARTICLE.apply(reqrepResGate))
dftest

```

9. On crée une colonne "Premier auteur" à partir de la colonne "creator » du sous-corpus OpenEdition

```

def take_first_author_when_creator(df):
    """
    Quand on a une colonne 'creator' cree une colonne 'Premier_auteur' avec le premier
    auteur uniquement
    """
    if 'creator' in df.columns:
        df['Premier_auteur'] = df['creator'].copy()
        idx_several_authors = df.Premier_auteur.isin("c(")
        df.loc[idx_several_authors, 'Premier_auteur'] = df.loc[idx_several_authors,
'Premier_auteur'].str.split('(')[1]
    else:
        print("No 'creator' column")

    return df

```

10. Scrapping et création de la colonne Researchgate dans le DataFrame df

```

def search_and_find_with_author(df, boolean_author = False):
    """
    Scrape et cree la colonne Researchgate dans la DataFrame df

    boolean_author permet de savoir si on recherche avec le nom de l'auteur
    """

    df = take_first_author_when_creator(df)

    if boolean_author:
        df['TITRE_ARTICLE_AUTEUR'] = df['TITRE_ARTICLE'].copy() + ' '
        df['TITRE_ARTICLE_AUTEUR'] = df['TITRE_ARTICLE_AUTEUR'] +
df['Premier_auteur'].str.split(',') [0]
        df = df.assign(Researchgate = df.TITRE_ARTICLE_AUTEUR.apply(reqrepResGate))

    df = df.assign(Researchgate = df.TITRE_ARTICLE.apply(reqrepResGate))

    return df

# test
dftest = dfArtGeo4.head(5)
dftest = dftest.assign(Researchgate = dftest.TITRE_ARTICLE.apply(reqrepResGate))

```

dfctest

	Code revue	REVUE	TITRE_ARTICLE	Auteur	date de mise en ligne	Researchgate
614	G_004	Autrepart	État développeur, État fragile : comment l'île...	Emmanuel Grégoire	07/12/2017	0
615	G_004	Autrepart	Vieillesse démographique en Chine : perspe...	Isabelle Attané	07/12/2017	0
616	G_004	Autrepart	Les mobilités quotidiennes, un révélateur des ...	Florence Boyer	07/12/2017	0
617	G_004	Autrepart	Emploi domestique et travail identitaire chez ...	Maud Laëthier	07/12/2017	0
618	G_004	Autrepart	Le registre péruvien de savoirs collectifs des...	Adriana Munoz Sanchez	07/12/2017	0

11. On crée maintenant les fichiers .csv avec la nouvelle colonne « ResearchGate »

```
def create_csv(dfArtGeoX, name_csv):
    """
    Creation des fichiers CSV avec la nouvelle colonne
    """
    dfArtGeoX.assign(Researchgate =
dfArtGeoX.TITRE_ARTICLE.apply(reqrepResGate)).to_csv(name_csv, encoding='utf-8')

    return 0

def create_csv_avec_loop(dfArtGeo, name_csv):
    """
    Meme chose mais avec une boucle
    """

    list_RG_responses = []
    bool_stop = False
    list_articles = dfArtEnv.TITRE_ARTICLE.values
    nb_articles = len(list_articles)
    csv_name_when_error = name_csv[:-4] + '_erreur.csv'
    # check articles already checked previously
    if os.path.isfile(csv_name_when_error):
        df_already_done = pd.read_csv(csv_name_when_error, index_col=0)
        # df_all = pd.DataFrame(list_articles, columns=['TITRE_ARTICLE'])
        # df_already_done = pd.merge(df_all, df_already_done, how = 'outer')

    else:
        # creation of the dataframe with the researchgate values as Not a Number (nan)
        df_already_done = pd.DataFrame(list_articles, columns=['TITRE_ARTICLE'])
        df_already_done['Researchgate'] = np.nan
        list_articles_done =
df_already_done[~df_already_done['Researchgate'].isnull()]['TITRE_ARTICLE'].values

    # shuffling the index in order to take an article randomly
    index_shuffle = np.random.permutation(len(df_already_done))[df_already_done.index]
    # loop
    for idx_art in index_shuffle:
        # article name of the row
        art_name = df_already_done.loc[idx_art, 'TITRE_ARTICLE']
        if not bool_stop:
            if art_name not in list_articles_done:
```

```

nb_articles))
    print('Article "%s..." : %d/%d' % (art_name[:15], idx_art,
try:
    list_RG_responses.append((art_name, reqrepResGate(art_name)))
    df_already_done.loc[idx_art, 'Researchgate'] = list_RG_responses[-
1][1]
except:
    print('Probleme, on ecrit un csv %s' % csv_name_when_error)
    # pd.DataFrame(list_RG_responses, columns = ['TITRE_ARTICLE',
'Researchgate']).to_csv(csv_name_when_error, encoding = 'utf-8')
    df_already_done.to_csv(csv_name_when_error, encoding = 'utf-8')
    bool_stop = True
else:
    print('Article "%s..." already done before' %art_name[:15])

if not bool_stop:
    df_art = pd.DataFrame(list_RG_responses, columns = ['TITRE_ARTICLE',
'Researchgate'])
    # dfArtGeo['Researchgate'] = df_art['Researchgate']
    dfArtGeo['Researchgate'] = df_already_done['Researchgate'].values
    dfArtGeo.to_csv(name_csv, encoding = 'utf-8')

return 0

```

12. On nomme les fichiers .csv en fonction de chaque revue traitée par le script

```

list_csv = [(dfArtGeo1, 'Annales de géographie.csv'),
(dfArtGeo3, 'Archéosciences.csv'),
(dfArtGeo4, 'Autrepart.csv'),
(dfArtGeo5, 'Cahiers de la recherche architecturale et urbaine.csv'),
(dfArtGeo6, 'Cahiers des amériques latines.csv'),
(dfArtGeo8, 'Cahiers dOutre-Mer.csv'),
(dfArtGeo9, 'Carnets de géographes.csv'),
(dfArtGeo10, 'Confins.csv'),
(dfArtGeo11, 'Cybergéo.csv'),
(dfArtGeo12, 'Echogéo.csv'),
(dfArtGeo13, 'Espace géographique.csv'),
(dfArtGeo14, 'Espace politique.csv'),
(dfArtGeo15, 'Espace populations sociétés.csv'),
(dfArtGeo16, 'Espaces et sociétés.csv'),
(dfArtGeo17, 'Espace-temps.csv'),
(dfArtGeo18, 'Etudes rurales.csv'),
(dfArtGeo19, 'Field Actions Science Reports.csv'),
(dfArtGeo20, 'Flux.csv'),
(dfArtGeo21, 'Géocarrefour.csv'),
(dfArtGeo22, 'Géographie et cultures.csv'),
(dfArtGeo23, 'Géographie économie société.csv'),
(dfArtGeo24, 'Géomorphologie.csv'),
(dfArtGeo25, 'Hérodote.csv'),
(dfArtGeo26, 'Histoire urbaine.csv'),
(dfArtGeo27, 'Hommes et migrations.csv'),
(dfArtGeo28, 'Linformation géographique.csv'),
(dfArtGeo29, 'Mappemonde.csv'),
(dfArtGeo30, 'Maghreb-Machrek.csv'),
(dfArtGeo31, 'Méditerranée.csv'),
(dfArtGeo32, 'Métropoles.csv'),
(dfArtGeo33, 'Monde Chinois.csv'),
(dfArtGeo34, 'Netcom.csv'),
(dfArtGeo35, 'Norois.csv'),

```

```

(dfArtGeo36, 'Outre-terre.csv'),
(dfArtGeo37, 'Photo-interpretation.csv'),
(dfArtGeo38, 'Physio-geo.csv'),
(dfArtGeo39, 'Pole sud.csv'),
(dfArtGeo40, 'Population.csv'),
(dfArtGeo41, 'Population et avenir.csv'),
(dfArtGeo42, 'Problèmes damérique latine.csv'),
(dfArtGeo43, 'Monde Chinois.csv'),
(dfArtGeo44, 'Quaternaire.csv'),
(dfArtGeo46, 'Revue de géographie alpine.csv'),
(dfArtGeo47, 'Revue déconomie régionale et urbaine.csv'),
(dfArtGeo48, 'Revue européenne des migrations internationales.csv'),
(dfArtGeo49, 'Revue géographique de lEst.csv'),
(dfArtGeo52, 'Rives méditerranéennes.csv')]
for dfArtEnv, name_csv in list_csv [14:] : ## changer par 'for dfArtEnv, name_csv in
list_csv[n:]:' pour skipper les n premières revues
    print('Revue %s'%name_csv)
    print("Nombre d'articles : %d"%len(dfArtEnv))
    # create_csv(dfArtEnv, name_csv)
    create_csv_avec_loop(dfArtEnv, name_csv)

```

13. Fusion des csv et gestion des erreurs

On obtient des erreurs aléatoires dues au serveur ResearchGate qui font qu'on ne peut pas obtenir le résultat attendu dans la colonne ResearchGate. Une solution brute-force est de lancer plusieurs fois le script et de recouper les articles qui n'ont pas pu avoir le résultat.

```

def gestion_erreurs(name_csv):
    """
    Change le csv 'name_csv' pour remplacer les valeurs dans les colonnes
    et le sauvegarde
    """
    dfArtGeo = pd.read_csv(name_csv, index_col=1)
    # on regarde les index des erreurs
    idxs_error = dfArtGeo[dfArtGeo.Researchgate == 2].index
    # on repasse ces erreurs par le processus de scrapping
    dfArtGeo_error = dfArtGeo.loc[idxs_error]
    dfArtGeo_error.assign(Researchgate =
dfArtGeo_error.TITRE_ARTICLE.apply(reqrepResGate))

    # on remet a jour le fichier csv initial avec les nouvelles valeurs
    dfArtGeo.at[idxs_error, 'Researchgate'] = dfArtGeo_error.Researchgate

    # on save
    name_csv_final = name_csv[:-4] + '_erreurs_traitees.csv'

```

```
dfArtGeo = pd.to_csv(name_csv_final)
return 0
for _, name_csv in list_csv:
    gestion_erreurs(name_csv)
```