



HAL
open science

L'usage de la plateforme HAL par des unités de recherche

Joachim Schöpfel

► **To cite this version:**

Joachim Schöpfel. L'usage de la plateforme HAL par des unités de recherche. I2D – Information, données & documents, 2020, 3 (3), pp.167-198. hal-03144160

HAL Id: hal-03144160

<https://hal.science/hal-03144160v1>

Submitted on 17 Feb 2021

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

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

L'usage de la plateforme HAL par des unités de recherche

Le cas de l'Université de Lille

The Usage of the HAL Platform by Research Laboratories – The Case of the University of Lille

Mots-clés

Archive ouverte, libre accès, science ouverte, unité de recherche

Keywords

Open access, open repository, open science, research laboratory

Résumé

L'article présente les résultats d'une étude sur l'usage de la plateforme HAL par les 62 laboratoires de recherche de l'Université de Lille, réalisée en avril 2020. L'analyse a porté sur les dépôts dans HAL, sur la création d'une collection et sur la part des documents en libre accès. La comparaison des unités selon leur domaine scientifique révèle quelques différences disciplinaires, aussi bien par rapport au nombre total des dépôts (notices et documents) que par rapport aux dépôts des publications en texte intégral. L'article distingue entre plusieurs approches types et propose quelques explications pour ces différences.

Abstract

The article presents the results of a study on the use of the HAL platform by the 62 research laboratories of the University of Lille, carried out in April 2020. The analysis focused on the deposits in HAL, on the creation of a collection and on the share of open access documents. The comparison of the units according to their scientific field reveals some disciplinary differences, both in relation to the total number of repositories (records and documents) and in relation to the repositories of full text publications. The article distinguishes between some typical approaches and offers explanations for these differences.

1. Introduction

La politique en faveur de la science ouverte définit la diffusion libre et directe des résultats de la recherche comme l'un de ses principaux objectifs. Ainsi, le Plan national pour la science ouverte du Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation préconise l'ouverture des publications scientifiques comme « pratique par défaut », soit par la publication dans des revues ou ouvrages nativement en accès ouvert, soit par dépôt dans une archive ouverte publique comme HAL dont le Ministère confirme le rôle national (MESRI 2018).

L'archive ouverte HAL¹ a été créée par le CNRS en 2001, sur le modèle du serveur de preprints arXiv. Administrée par le CCSD², la plateforme HAL est devenue, au fil des ans et « dans les faits, (...) un outil commun et ouvert à l'ensemble des chercheurs, même s'il existe des archives institutionnelles développées en-dehors de HAL » (Magron 2017). Au cœur de l'écosystème de la science ouverte en France, HAL contient aujourd'hui (juin 2020) plus de 730 000 documents scientifiques, plus 1,3 million de notices bibliographiques.

¹ HAL archive ouverte <https://hal.archives-ouvertes.fr/>

² Centre pour la Communication Scientifique Directe <https://www.ccsd.cnrs.fr/>

Son système d'information est centralisé mais, dans la mesure où il permet la création de sous-ensembles, portails ou collections, HAL peut être considérée comme une archive mutualisée qui réalise une économie d'échelle significative pour l'ensemble des organismes et établissements de l'enseignement supérieur et de la recherche (ESR) français : « La tâche de maintenance des ordinateurs, avec une équipe compétente, est transférée à une unité commune, bien dotée en ressources grâce à un effet d'échelle. Les chercheurs ou les bibliothécaires conservent la précieuse tâche de préparer, de mettre en forme et d'afficher les résultats de leur propre institution » (Baruch 2007). Ainsi, HAL représente un modèle unique et exemplaire dans un paysage international caractérisé par une multitude d'archives ouvertes éclatées et disparates (Arlitsch & Grant 2018).

HAL a été développée comme une plateforme de la « voie verte » du libre accès à l'information scientifique, c'est-à-dire un serveur où les chercheurs déposent eux-mêmes leurs publications pour une diffusion libre, accessible à tous (Harnad et al. 2004). Le principe de l'auto-archivage place les pratiques des chercheurs au centre des enjeux : comment acceptent-ils ce nouveau mode de communication ? Que déposent-ils ? Quelles sont leurs motivations et leurs attitudes à l'égard du libre accès et des archives ouvertes ? Quelle est la relation avec des variables indépendantes démographiques (âge, sexe, pays...), professionnelles (statut, carrière, discipline...), institutionnelles (type d'organisation, politique institutionnelle...) et techniques (ergonomie des fonctionnalités, qualité des services...) ?

De leur côté, les établissements et organismes de l'ESR ont compris l'intérêt institutionnel de ces plateformes, pour augmenter la visibilité et l'impact de la production scientifique d'une institution, pour l'évaluation des structures et personnels et pour le monitoring de leurs politiques et investissements scientifiques (COAR 2015). La visibilité est comprise ici comme la présence, le référencement et l'accessibilité (avec métadonnées) des articles, chapitres, livres, thèses, rapports, communications etc. des chercheurs d'une institutions sur le Web, avec leur identification et une attribution claire et sans équivoque de ces résultats à cette institution, notamment via les affiliations des auteurs mais aussi (cf. plus loin) par leur regroupement sous forme d'une collection ou d'un portail. Cette nouvelle approche des archives ouvertes implique au moins quatre changements :

- La mise en place d'autres vecteurs d'alimentation que l'auto-archivage ;
- La création de nouvelles fonctionnalités ;
- L'importance de la qualité et de l'exhaustivité des métadonnées (notices) ;
- Le développement de l'interopérabilité avec d'autres systèmes (plateformes de revues, système de gestion recherche...).

Dans cet écosystème en évolution rapide, nous nous intéressons au rôle des laboratoires universitaires, c'est-à-dire des unités ou centres de recherche au sein d'une université, créés par délibération du conseil d'administration de l'université après avis du conseil académique, pouvant administrer les dotations globales de fonctionnement et d'équipement qui leur sont allouées par les organes directeurs de l'université et faisant l'objet d'une évaluation par le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCÉres)³. Dans la mesure où il s'agit de la « structure sociale constituée donnant un cadre de travail aux chercheurs »⁴, nous nous interrogeons sur la stratégie institutionnelle de cette structure par rapport à HAL comme dispositif central et national de la voie verte du libre accès en France.

³ Cf. Code de l'Éducation L713-1 et Code de la Recherche L112-1, L114-3-1, L321-3

⁴ Wikipédia https://fr.wikipedia.org/wiki/Laboratoire_de_recherche 18 juin 2020

La question est double : comment les unités se servent-elles de ce dispositif pour rendre leur production visible et en augmenter l'impact, pour assurer un suivi de cette production et pour contribuer à l'évaluation par le HCéres par l'édition des listes de publications ; et comment ces unités soutiennent ou incitent-elles leurs chercheurs à déposer leurs publications dans HAL. L'objectif de cette étude est de contribuer à la compréhension de l'écosystème de la science ouverte sous l'aspect organisationnel et de dispositif numérique (Larroche 2018), sur le terrain universitaire.

Pourquoi les laboratoires ? La plupart des études empiriques sur le développement de la science ouverte en général et du libre accès aux publications en particulier mettent le focus sur les individus (chercheurs, auteurs) ou sur les établissements (universités) mais négligent le niveau intermédiaire, les laboratoires de recherche et leurs directions et administrations. Or, leur rôle est essentiel, aussi bien comme cadre de référence immédiat des chercheurs, comme une sorte de courroie de transmission des opinions et des demandes des chercheurs, que comme représentants de l'établissement et/ou l'organisme de recherche et « gardiens » de l'application des décisions et des règles institutionnelles sur le terrain et comme unité d'évaluation (Schöpfel et al. 2016). C'est pour cette raison que nous nous intéressons aux unités de recherche comme un élément clé de la médiation et du développement de la science ouverte.

Les variables dépendantes de notre étude sont en particulier :

- La présence d'une collection HAL.
- La présence d'un identifiant institutionnel HAL validé.
- Le nombre total des dépôts sur HAL.
- La part des documents par rapport au nombre total des dépôts (« degré d'ouverture »).

Comme variables indépendantes, l'analyse fait le lien avec le domaine scientifique et la taille des unités, en termes de ressources humaines (nombre de chercheurs et d'enseignants-chercheurs). Il s'agit d'une étude explorative et préliminaire dans le cadre d'un projet de recherche plus large, avec les laboratoires des universités du programme IDEX.

Le plan de l'article : La section 2 fait un état des connaissances sur le dispositif HAL, ses collections et ses usages, et sur les pratiques des communautés scientifiques par rapport à HAL. La section 3 décrit la méthodologie de notre étude. La section 4 présente les résultats, les collections des laboratoires, le nombre de dépôts, la part des documents en libre accès, tout cela par rapport aux domaines et disciplines scientifiques. La discussion de la section 5 porte sur les limitations méthodologique de cette étude, sur la possibilité de regrouper les unités en clusters, sur les variables pouvant expliquer ces résultats, et sur l'évolution des résultats depuis 2019. La conclusion (section 6) indiquent plusieurs pistes pour aller plus loin dans l'analyse des pratiques des laboratoires par rapport au dispositif HAL.

2. Le contexte de l'étude

Selon une enquête récente, la quasi-totalité des établissements et organismes de l'ESR français propose aujourd'hui une solution de dépôt et/ou de signalement à leurs chercheurs qui est, pour 80%, la plateforme HAL ; neuf universités sur dix ont fait le choix d'un portail institutionnel sur HAL (Couperin 2020). Les autres ont établi une connexion avec HAL, ou le feront. Les objectifs de HAL et les solutions locales sont différents mais complémentaires. « La mission de HAL est avant tout une mission de promotion et d'affichage de la production scientifique de la recherche française dans le monde » tandis qu'une archive ouverte institutionnelle sert avant tout à la valorisation de la production scientifique de l'établissement (Lafon 2013a).

La préférence des universités pour la voie verte du libre accès à l'information scientifique a été confirmée par une enquête internationale, à partir d'un corpus de 4,6 millions d'articles de revues

publiés par des chercheurs de 963 universités (Robinson-Garcia et al. 2020). Selon cette étude, 41% articles sont librement accessibles sur Internet, et de ces presque 1,9 millions articles, 77% ont été déposés dans des archives ouvertes et peuvent y être lus et téléchargés. Par rapport au choix de la voie verte par les universités, la France figure parmi les tout premiers grands pays, devant les Etats Unis, le Brésil et l'Allemagne.

2.1. Les collections HAL

HAL, comme mentionné plus haut, a été conçue comme un système centralisé, qui peut paraître allant « contre la doctrine du distribué, du local, des dépôts » (Baruch 2007). Néanmoins, grâce à la création de sous-ensembles, « le dilemme entre les archives locales et les archives centrales ou thématiques disparaît » (idem). En 2020, HAL compte plus de 130 portails institutionnels (organismes de recherche, universités, COMUE, grandes écoles...) et plus de 2 500 collections qui correspondent à un « regroupement d'un ensemble d'articles sous une URL particulière (...) et pouvant être personnalisée par le propriétaire de la collection » (Charnay & Michau 2007). La personnalisation concerne le graphisme, les textes de la page d'accueil de la collection et l'intégration de plug-ins (statistiques, actualités...). Une collection se constitue en apposant sur des dépôts le tampon de la collection, ce qui peut se faire manuellement ou automatiquement sur un critère spécifique (affiliation, domaine...). On trouve ainsi des collections d'événements et de projets scientifiques mais surtout, des collections de laboratoires, servant de « présentoir » de leurs publications.

Parmi ces collections, on trouve également des universités, écoles et COMUES qui ont préféré de créer une collection « sans passer le cap de la création d'un Portail qui engage davantage l'institution dans le cadre d'une convention et rend nécessaire la mobilisation de ressources pour l'administrer (...) Ces collections peuvent être également un regroupement destiné à l'export des références vers leur propre archive institutionnelle locale » (Magron 2017). De même, certains laboratoires ont fait le choix de créer un portail. En 2017, 3% des portails HAL étaient en réalité des portails de laboratoires (CNRS 2018), comme par exemple celui créé en 2016 par le Laboratoire d'analyse et d'architecture des systèmes (LAAS) de Toulouse qui contient plus de 5 000 documents répartis en 40 collections de départements, d'équipes, de services et de conférences⁵.

La différence entre portails et collections HAL est avant tout une question d'administration et d'étendu des fonctionnalités, avec des cibles différents⁶ :

- Un portail est une fonctionnalité de HAL qui permet à toute institution de gérer la littérature scientifique produite en son sein. Est liée à l'institution toute publication dont au moins un auteur est affilié à une structure de recherche qui a pour tutelle cette institution. Un portail HAL a un nom de domaine spécifique, une interface de dépôt et une charte graphique personnalisée. Il est possible d'ajouter des métadonnées spécifiques (disciplines, types de document, métadonnées de description du document) autres que celles de HAL. L'administrateur d'un portail peut gérer des référentiels (auteurs, unités de recherche, revues, projets ANR etc.), les dépôts, les doublons et les utilisateurs.
- Une collection est constituée d'un ensemble de dépôts qui sont sélectionnés à partir de critères choisis par le gestionnaire de collection. Chaque dépôt répondant au critère choisi est « tamponné » manuellement ou automatiquement, et la sélection ainsi constituée apparaît dans une extraction automatique de la base par tampon, indépendamment de tous les autres documents contenus dans HAL. Tout utilisateur ayant un compte dans HAL peut demander la création d'une collection personnelle à son nom pour constituer une liste des articles qui

⁵ Archive ouverte HAL-LAAS <https://hal.laas.fr/>

⁶ Documentation HAL <https://doc.archives-ouvertes.fr/>

l'intéressent. Plus officiellement, un responsable de laboratoire ou d'établissement peut demander la création d'un tampon avec un nom institutionnel, afin de construire facilement une liste visible d'articles en mettant un lien vers cette liste depuis le site internet officiel de son unité. Une collection peut également être utile pour regrouper des actes de congrès. La présentation d'une collection est personnalisable (page d'accueil, choix d'onglets prédéfinis pour le menu) et a une adresse url stable. Mais il n'y a ni métadonnées spécifiques, gestion de référentiel, de dépôts etc.

Tandis que seulement une structure institutionnelle peut demander la mise en place d'un portail, tout utilisateur avec un compte HAL (chercheur, responsable d'une unité, chef de projet, documentaliste...) peut demander la création d'une collection. Pour cette raison, le gestionnaire d'une collection, contrairement au portail, n'a aucun droit sur les dépôts constituant sa collection, ce droit restant acquis aux propriétaires ou co-propriétaires du dépôt, ainsi qu'aux référents des laboratoires présents dans le dépôt. La gestion de la collection n'autorise donc pas le gestionnaire à intervenir sur les dépôts même de sa collection ; les dépôts sont marqués pour être 'visibles' dans une collection donnée, mais leur appartenance à une collection n'implique pas la propriété du dépôt pour le gestionnaire de la collection.

L'identification d'une unité de recherche sur HAL est facilitée par un référentiel de structures (institutions, laboratoires, équipes, regroupements...) ⁷ qui, au niveau d'un laboratoire, contient le nom du laboratoire, son sigle, l'ensemble des institutions, des universités, et des grandes écoles auquel il est rattaché, l'URL du laboratoire et son adresse, avec une information indiquant si le laboratoire est en activité ou s'il est fermé, et une correspondance avec le Référentiel national des structures de recherche ⁸ (Charnay & Michau 2007). La qualité de ce référentiel reste, cependant, un enjeu, notamment pour la validation des structures (CNRS 2018).

2.2. *Les choix institutionnels*

Quelles sont les motivations pour un laboratoire de créer son identifiant et/ou une collection sur HAL ? Magron (2017, p.8) énumère sept objectifs potentiels :

- La mise à disposition en accès libre et gratuit de la production scientifique ;
- La réalisation d'une vitrine de la recherche et de la diversité des productions de l'établissement ;
- L'alimentation des pages web des laboratoires et des chercheurs ;
- Le recensement de la production de ses communautés ;
- Le partage de connaissances ;
- L'analyse de la production scientifique et technique de l'établissement, y compris par la production d'indicateurs ;
- La préservation à terme des documents produits.

L'environnement institutionnel joue un rôle important, en particulier la politique d'établissement par rapport à la science ouverte. L'enquête du consortium Couperin (2020) a révélé que plus de la moitié des organismes de recherche et universités ont intégré l'accès libre aux publications dans leur politique générale d'établissement mais que seulement 23% demandent explicitement le signalement de toutes les publications dans l'archive institutionnelle, et seulement 15% exigent leur dépôt en texte intégral.

⁷ Accès Unifié aux Référentiels HAL AURéHAL <https://aurehal.archives-ouvertes.fr/>

⁸ RNSR <https://appliweb.dgri.education.fr/rnsr/>

D'autres choix d'établissement concernent l'utilisation des données et indicateurs de HAL pour le pilotage de la recherche et les moyens dévolus au libre accès en général et à l'administration, à l'alimentation et à la promotion de HAL en particulier (CNRS 2018).

Ce choix institutionnel doit être mis en rapport avec la politique du Ministère en faveur de la science ouverte en général et du libre accès aux publications des chercheurs en particulier (cf. Chartron & Schöpfel 2017), relayée par les organismes de recherche et la Conférence des présidents d'université qui, dans un contexte « marqué par la volonté de développer la science ouverte portée par le plan national pour la science ouverte de 2018 », a proposé le 14 mai 2020 que les différentes instances d'évaluation internes ou externes aux établissements prennent désormais en compte dans l'évaluation des enseignants-chercheurs la publication des articles en accès ouvert qu'elles s'appuient pour cela sur la liste des publications déposées dans une archive ouverte, soit directement dans l'archive ouverte nationale HAL, soit dans une archive institutionnelle⁹. Par ailleurs, la CPU va rejoindre le nouveau comité de pilotage de HAL, à partir de 2021¹⁰.

2.3. *Le rôle des laboratoires*

« Briques de base »¹¹ de la recherche dans les établissements de l'ESR, les unités de recherche jouent plusieurs rôles par rapport aux archives institutionnelles en général et HAL en particulier :

- **Soutien et incitation** : Selon une enquête auprès de 432 directeurs d'unités du CNRS en 2014, beaucoup avaient un avis favorable du libre accès et soutenaient la plateforme HAL, y compris pour l'auto-archivage et le signalement de la production du laboratoire. En revanche, moins de 40% se prononçaient en faveur d'une obligation de dépôt sur HAL (Schöpfel et al. 2016).
- **Information et formation** : Au sein des universités, le rôle des laboratoires pour informer les chercheurs sur les archives ouvertes et HAL est réel quoique moins important que l'impact du SCD et de la Direction Recherche (Couperin 2020). Néanmoins, une enquête à Toulouse a souligné le rôle à jouer par les directions et les conseils de laboratoire, pour la promotion de HAL et l'incitation au dépôt (Lafon 2013b). Le cas d'une formation interne de l'UMR LAAS de Toulouse est exemplaire (Smith 2017) ; mais en l'absence d'un recensement de toutes ces actions de formation, il paraît peu probable qu'il y ait ce genre d'actions dans beaucoup de laboratoires, notamment dans les unités de recherche universitaires (équipes d'accueil).
- **Dépôt et signalement** : Si en règle générale les chercheurs déposent eux-mêmes les fichiers de leurs publications sur HAL (cf. Gandossi 2011, Gayoso 2020), les documentalistes, bibliothécaires et d'autres personnels rattachés aux laboratoires, interviennent pour le signalement des publications, par le biais de la création ou de l'import de notices. « Le signalement est opéré pour partie par les professionnels de l'IST et vise à minima à identifier la production de l'établissement » (Magron 2017). L'analyse de presque 60 000 dépôts en sciences de la vie (dont 76% sans documents) a ainsi montré que 86% des notices avaient été créées par des intermédiaires institutionnels (Prime-Claverie & Mahé 2013). En SHS, cette contribution a été évaluée à plus de 40%, notamment en droit, économie et gestion où elle est supérieure à 50% (Prime-Claverie & Mahé 2017a). Une étude récente, sur un corpus de 368 revues de cinq disciplines SHS, a estimé la part des dépôts par auto-archivage (chercheurs) à

⁹ CPU Évaluation des enseignants-chercheurs et science ouverte - Propositions de recommandations <http://www.cpu.fr/actualite/evaluation-des-ec-et-science-ouverte-propositions-de-recommandations/>

¹⁰ CPU HAL : une gouvernance partagée et un modèle de financement pérenne <http://www.cpu.fr/actualite/hal-une-gouvernance-partagee-et-un-modele-de-financement-perenne/>

¹¹ CNRS La recherche <http://www.cnrs.fr/la-recherche>

38% ; le reste correspond aux imports en provenance d'archives locales et à la création de notices par les documentalistes et bibliothécaires dans les SCD et laboratoires (Gayoso 2020).

2.4. *Les facteurs de succès*

Magron (2017) met en avant qu'une « politique d'incitation et d'animation autour du dépôt au sein d'une équipe de recherche est facteur de succès », et elle cite le cas de deux laboratoires (Triangle et CRESSON) qui ont mis en place « un accompagnement personnalisé des chercheurs (...) par les professionnelles de l'information, soutenues par la direction des laboratoires ».

Même si la plupart des chercheurs dans notre domaine des sciences de l'information et de la communication semble acquis au principe du libre accès, l'acceptation d'une collection sur HAL, de l'auto-archivage et du signalement des publications ne va pas de soi. Notre étude de la collection HAL du laboratoire GERiCO a montré plusieurs variables qui semblent importantes voire décisives pour mener à bien l'initiative d'une collection de laboratoire sur HAL (Schöpfel et al. 2018) :

- Le soutien institutionnel par la Direction Recherche et par l'équipe de direction du laboratoire, notamment pour la communication et promotion auprès de la communauté scientifique.
- Le soutien par le SCD, en termes de gestion (administration), de communication et formation, de conseil et d'assistance technique et de liaison avec l'équipe du CCSD, mais aussi en termes d'aide au signalement par la création de notices sur la plateforme HAL.
- L'échange direct avec l'équipe de HAL et d'autres membres du réseau des utilisateurs de HAL¹² et l'évolution des fonctionnalités de la plateforme.
- Le débat au sein du laboratoire, par messagerie, en assemblée générale, séminaire etc.

2.5. *Les différences disciplinaires*

Le baromètre européen de la science ouverte montre très clairement des différences disciplinaires par rapport au libre accès en général et à la voie verte en particulier, où on constate davantage de dépôts en sciences de la vie et santé, physique, mathématiques et astronomie qu'en sciences humaines et sociales, à l'exception de la psychologie¹³. Ces chiffres montrent aussi que les différences disciplinaires dépendent de la politique et du développement particulier du libre accès dans chaque pays.

Les différences disciplinaires des dépôts dans HAL ont été étudiées dès le lancement de la plateforme. Ainsi, Baruch (2007) a fait remarquer qu'en physique et en mathématiques, « l'initiative d'ArXiv a fait de l'auto-archivage la norme, en particulier en physique des hautes énergies ; cependant, d'autres domaines ne suivent pas le même schéma » (cf. aussi Wojciechowska 2017). Dix ans plus tard, les mêmes différences persistent ; ainsi, Mahé (2017) observe qu'on retrouve dans les archives ouvertes « de grandes variations disciplinaires, les domaines les plus propices étant la physique, les mathématiques et l'informatique ». Aussi, les directeurs d'unités du CNRS en mathématiques, physique et informatique soutiennent davantage HAL que dans d'autres domaines (Schöpfel et al. 2016).

Une raison pour ces différences est probablement, pour certaines disciplines, l'interconnexion de HAL avec des plateformes communautaires reconnues, comme arXiv pour les physiciens et RePEc pour les économistes : « La discipline Economie et finances est la discipline qui est représentée par le plus grand nombre de dépôts dans HAL-SHS, et ce, quasiment depuis le début de l'archive (...) L'une (des raisons) est vraisemblablement l'interconnexion de HAL-SHS avec RePEc (Research Papers in Economics), la base internationale de référence dans cette discipline. A partir de 2007 en effet, tout dépôt en

¹² CasuHal <https://iww.inria.fr/casuhal/>

¹³ EU Open Science Monitor https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/trends-open-access-publications_en

économie dans HAL-SHS est reversé dans RePEc, augmentant ainsi la visibilité internationale de la publication » (Magron 2017).

L'existence de logiques disciplinaires de participation et de dépôt devient visible et se révèle à plusieurs niveaux : le nombre total et la représentativité des dépôts par rapport à la production réelle, le type de ressources déposées (par exemple, preprints en économie, communications en informatique...), le rapport entre dépôts de documents et notices sans documents. La part des documents en libre accès est plus élevée en SHS (31%) qu'en sciences de la vie (24%). L'enquête de Mahé & Prime-Claverie (2017a, cf. aussi Prime-Claverie et Mahé 2013) témoigne d'une forte mobilisation des chercheurs en SHS : « plus de 90 % des contributeurs de HAL-SHS sont des chercheurs (...) ; ils participent à plus de 57 % des dépôts sur HAL-SHS, contre 26 % en sciences de la vie ».

2.6. *Stratégies individuelles et collectives*

La plupart des études sur l'usage des archives ouvertes mettent le focus sur les pratiques individuelles. Ainsi, Mahé & Prime-Claverie (2017b) ont identifié trois groupes de chercheurs en sciences humaines et sociales, avec des variations disciplinaires : ceux qui déposent ponctuellement sur HAL (48%), ceux qui déposent régulièrement (35%) et ceux qui déposent de temps en temps, d'une manière irrégulière (9%). Mais le lien avec les stratégies institutionnelles reste incertain. Rappelons aussi, par exemple, la préférence des chercheurs en SHS pour les réseaux sociaux académiques, notamment Academia, quand il s'agit de la diffusion de leurs publications (Gayoso 2020).

Pour les sciences de la vie, Prime-Claverie & Mahé (2013) ont montré que moins de la moitié des dépôts de documents est effectuée par les auteurs eux-mêmes. En ce qui concerne les intermédiaires de l'institution, leur analyse révèle une certaine irrégularité des dépôts qu'elles expliquent par deux raisons : « il pourrait s'agir d'une mise à jour du recensement de la production scientifique dans le cadre (...) de campagnes d'évaluation (tous les quatre ans environ) ou d'un 'effet de rattrapage' par le dépôt massif d'enregistrements de nouveaux entrants (institutions qui ouvrent un site de dépôt) ». Mais elles admettent que des analyses plus approfondies sur les motivations des déposants sont nécessaires pour mieux comprendre les logiques à l'œuvre.

Une étude sur le campus scientifique de l'Université de Toulouse 3, avec 65 unités de recherche en sciences, technologies et médecine a révélé des pratiques multiples au sein des laboratoires, voire pour certains laboratoire l'absence quasi-totale de pratiques. Douze laboratoires (18%) avaient une collection HAL, avec une alimentation manuelle ou automatique, et sept laboratoires (11%) affichaient une exhaustivité de signalement sur HAL d'au moins 80%, par rapport au signalement dans le Web of Science (Lafon 2013b).

Nous avons mené une première analyse en 2019 à partir des dépôts sur HAL des unités de recherche de l'Université de Lille, pour évaluer l'utilisation de HAL par ces laboratoires. A partir de plusieurs éléments – présence d'une collection HAL, nombre de dépôts, taux de documents – nous avons identifié trois catégories qui ne recouvrent pas l'ensemble des unités (Schöpfel et al. 2019) :

- 9% des unités de recherche utilisent HAL comme une archive institutionnelle, comme vitrine pour leur production scientifique, avec une collection qui regroupe beaucoup de publications et un taux plus élevé de documents en accès libre.
- 29% des unités de recherche utilisent HAL comme une sorte de gestionnaire de références bibliographique, avec un nombre important de notices mais peu de documents ; la finalité est probablement de pouvoir produire des listes pour l'évaluation.

- 58% des unités ne semblent pas avoir de stratégie par rapport à HAL, pas de collection, peu de dépôts. Ce qui ne veut pas dire qu'il n'y ait pas de dépôts et/ou de signalement ailleurs, sur le site du laboratoire, sur une autre plateforme etc.

2.7. Objectif

Notre nouvelle étude met à jour l'analyse de 2019. L'objectif est d'étudier l'utilisation de HAL par les laboratoires de recherche de l'Université de Lille, comme un échantillon représentatif pour les laboratoires universitaires de l'ESR français. L'étude porte surtout sur le nombre de dépôts, sur la part des documents en libre accès et sur la création d'une collection, et essaie de dégager les différences disciplinaires.

3. Méthodologie

Nous avons constitué un échantillon exhaustif avec les 62 unités de recherche de l'Université de Lille à partir du site de l'université et d'une liste interne de la Direction générale déléguée recherche, valorisation et ingénierie de projets de l'Université de Lille, d'avril 2020.

Pour chaque unité, nous avons déterminé le nom, l'acronyme et/ou le numéro, le statut (structure universitaire ou gouvernance mixte avec des organismes de recherche), le nombre de chercheurs, le domaine de recherche¹⁴ et la discipline¹⁵ ; tout cela a été fait à partir des sites web de l'université et des unités et complété avec les informations du RNSR et du moteur de la recherche et de l'innovation scanR¹⁶.

Dans un deuxième temps, nous avons déterminé pour chaque unité la présence d'un identifiant institutionnel et/ou d'une collection sur la plateforme HAL.

Dans un troisième temps, nous avons évalué pour chaque unité le nombre de documents déposés (documents avec texte intégral) et de notices (métadonnées sans documents) par une recherche directe sur HAL.

A partir de ces chiffres, nous avons effectué des analyses statistiques en fonction de plusieurs variables indépendantes (statut, domaine, taille). Nous avons essayé de regrouper les unités en « clusters ». Nous avons également comparé ces chiffres avec les résultats de 2019 pour décrire l'évolution des dépôts, en tenant compte du fait d'une restructuration des unités de recherche de Lille début 2020.

Les analyses ont été faites en avril 2020. Les tests statistiques (χ^2) ont été réalisés sur un niveau $p=0.05$.

4. Résultats

4.1. Statut, domaine et discipline

23 laboratoires (37%) sont des unités propres de l'Université de Lille, financées (budget, équipement, personnel) uniquement par l'université ; 32 laboratoires (52%) sont des unités mixtes, situées sur le campus de l'Université de Lille mais partiellement financées par d'autres organismes, dont le CNRS et l'INRIA ; 7 unités de recherche (11%) relèvent de l'INSERM. La répartition par domaine scientifique montre l'importance des domaines sciences et technologies (ST) et des sciences de la vie et de la santé (SDV santé) qui représentent plus de deux tiers des laboratoires de Lille (N=43, tableau 1) :

Sciences et technologies	ST	22
--------------------------	----	----

¹⁴ Sciences et technologies (ST) ; sciences de la vie et de la santé (SdV santé) ; arts, lettres, langues, sciences humaines et sociales (ALLSHS) ; droit, économie, gestion (DEG).

¹⁵ Nomenclature du Répertoire national des structures de recherche (RNSR) <https://appliweb.dgri.education.fr/rnsr/>

¹⁶ ScanR <https://scanr.enseignementsup-recherche.gouv.fr/>

Sciences de la vie et de la santé	SdV santé	21
Arts, lettres, langues, sciences humaines et sociales	ALLSHS	12
Droit, économie, gestion	DEG	7

Tableau 1. Répartition des unités par domaine scientifique (N=62)

Quant aux domaines des sciences humaines et sociales (ALLSHS) et au droit, à l'économie et à la gestion (DEG), ils représentent environ un tiers des laboratoires (N=19).

Les unités de recherche de Lille couvrent l'ensemble des disciplines scientifiques ; cependant, la répartition par discipline confirme l'importance de la recherche en médecine et en biologie (une unité peut relever de plusieurs disciplines ; tableau 2).

1	Mathématiques et leurs interactions	3
2	Physique	6
3	Sciences de la terre et de l'univers, espace	8
4	Chimie	13
5	Biologie, médecine et santé	24
6	Sciences humaines et humanités	13
7	Sciences de la société	11
8	Sciences pour l'ingénieur	12
9	Sciences et technologies de l'information et de la communication	7
10	Sciences agronomiques et écologiques	6

Tableau 2. Répartition des unités par discipline (N=62)

26 unités sont répertoriées avec deux ou davantage de disciplines (42%).

Le nombre médian des chercheurs dans les unités de recherche est 36. Ce chiffre correspond aux unités en ST, il est supérieur en ALLSHS (61) et DEG (53) mais inférieur en SdV santé (20).

4.2. Identifiant et collection HAL

Toutes les unités de recherche de l'Université de Lille disposent d'au moins un identifiant HAL dans le référentiel des structures de recherche¹⁷. Il est donc possible de les identifier sur HAL, lancer des recherches et extraire des données via les API.

19 unités (31%) ont créé une collection de laboratoire sur HAL, pour rendre leur production scientifique visible. Les différences par domaine sont significatives (figure 1).

¹⁷ Cf. Accès Unifié aux Référentiels HAL AURÉHAL <https://aurehal.archives-ouvertes.fr/>

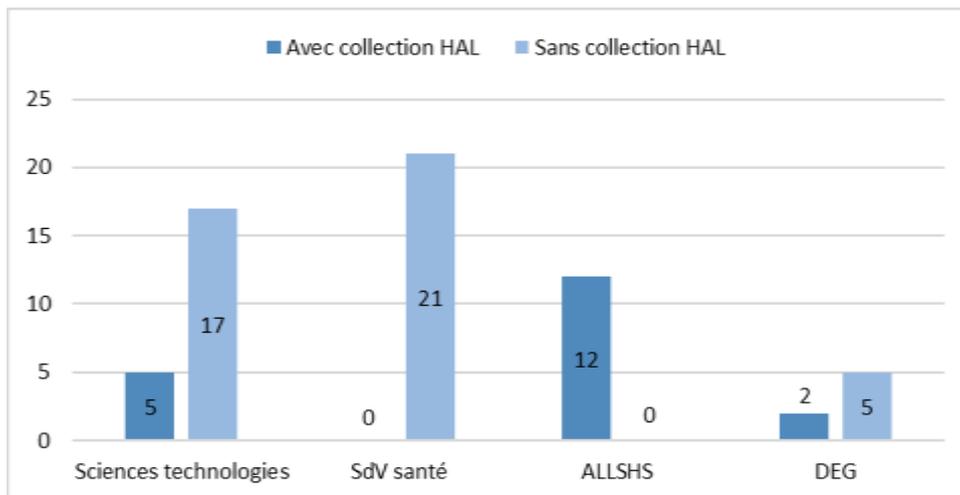


Figure 1. Collection HAL par domaine scientifique (N=62)

Ce qui fait la différence, c'est surtout l'absence de collections issues des laboratoires en SdV santé, et le fait que la totalité des unités en ALLSHS ont créé leur collection sur HAL. La situation dans les autres domaines est plus mitigée ; un quart environ des unités ST et DEG ont une collection sur HAL, les autres pas. Notamment il n'y a que peu de collections dans les sciences pour l'ingénieur et en sciences agronomiques et écologiques mais la différence avec les autres disciplines ST n'est pas statistiquement significative.

4.3. Le nombre de dépôts

Les 62 unités de recherche de l'Université de Lille totalisent 50 266 dépôts sur HAL (avril 2020). La moyenne de 811 dépôts par unités est trompeuse car la moitié des unités ont moins de 300 dépôts sur la plateforme (médiane = 314). Figure 2 montre une distribution inégale, avec une longue traîne d'unités avec peu de dépôts.

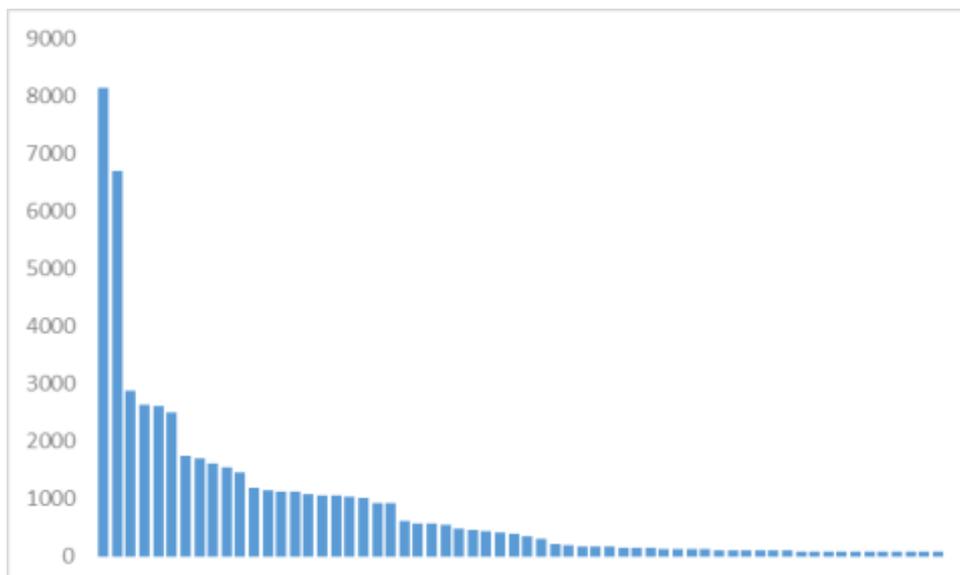


Figure 2. Nombre de dépôts sur HAL pour les 62 unités (N=50 266)

- Dans ce ranking, le premier laboratoire, IEMN (*Institut d'électronique, de microélectronique et de nanotechnologie*), totalise plus de 8 000 dépôts sur HAL, suivi par CRISTAL (*Centre de recherche en informatique, signal et automatique*) avec presque 7 000 dépôts ;

- Après, on peut distinguer plusieurs paliers : un groupe de quatre unités avec 2-3 000 dépôts ;
- Puis cinq unités avec 1-2 000 dépôts ;
- Puis onze unités avec autour de 1 000 dépôts ;
- Puis toutes les autres, avec 500 dépôts ou moins.

Il ne s'agit pas d'une distribution Pareto au sens strict (80/20) mais d'une longue traîne, où 30% des unités « produisent » 80% des dépôts sur HAL, où à l'autre bout de l'échelle, 82% des unités produisent 34% des dépôts, et où il n'y a aucune unité sans dépôt.

La corrélation entre le nombre de chercheurs et le nombre de dépôts sur HAL est haute, avec un coefficient $r=0.82$ (figure 3).

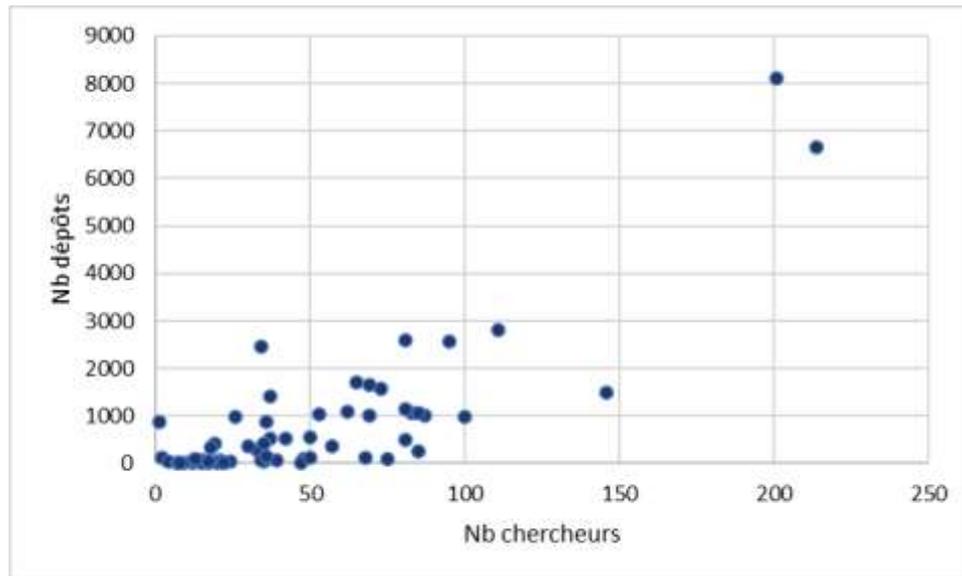


Figure 3. Nombre de dépôts sur HAL et nombre de chercheurs pour chaque unité (N=62)

L'analyse de cette distribution par domaine révèle qu'il y a surtout deux particularités significatives : une médiane relativement élevée de dépôts en sciences et technologies et dans les domaines de droit, économie et gestion ; et l'absence de dépôts en sciences de la vie et de la santé, notamment en biologie et pharmacologie (tableau 3).

Domaine	Nb unités	Nb dépôts	Nb dépôts par unité (médiane)
ST	22	30484	868
SdV santé	21	2561	66
ALLSHS	12	11386	765
DEG	7	5835	999

Tableau 3. Répartition des dépôts par domaine scientifique (N=50 266)

Quand on regarde ces chiffres sous l'aspect des collections, il s'avère que 31 884 dépôts (63%) se trouvent au sein d'une collection, même si ces collections ne représentent que 31% des unités. La différence entre les domaines scientifiques est significative (figure 4).

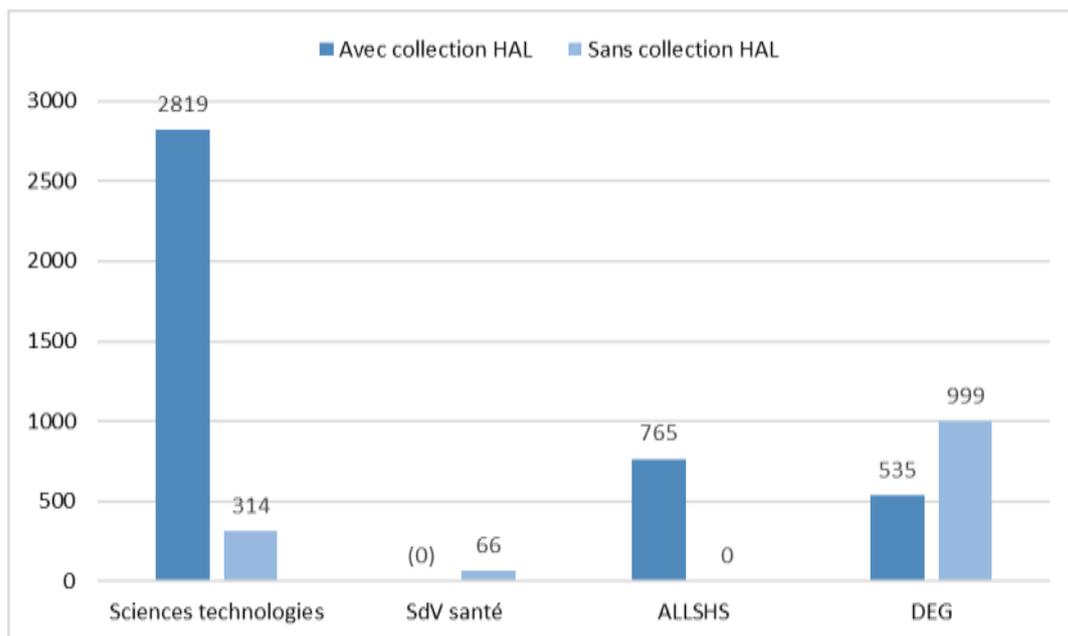


Figure 4. Médiane des dépôts par domaine scientifique, avec ou sans collection (N=50 266)

L'effet d'une collection est surtout visible en sciences et technologies où la médiane des dépôts par unité est presque dix fois plus élevée que sans collection. On voit très clairement le rôle que jouent un petit groupe de laboratoires « moteur » comme l'IEMN et CRISTAL mais aussi, depuis 2019, l'UCCS (*Unité de catalyse et de chimie du solide*).

La situation en DEG semble différente, puisque la médiane des unités sans collection est plus élevée que celle des unités avec collection. Certes, à cause du petit nombre d'unités en DEG (seulement sept), il ne faut sans doute pas trop interpréter ou généraliser mais regarder ces laboratoires et leurs spécificités de près pour comprendre l'effet des choix individuels et collectifs.

4.4. Le degré d'ouverture

Quelle est la part des documents (plein texte), combien de dépôts sont accompagnés du fichier de la publication (article, chapitre, communication, thèse...)? D'une manière générale, la part des documents (N=11 506) par rapport à l'ensemble des dépôts des unités de Lille s'élève à 23% ; ce pourcentage est légèrement plus bas que pour l'ensemble des 2,7 million dépôts sur la plateforme HAL (25%, cf. figure 5).

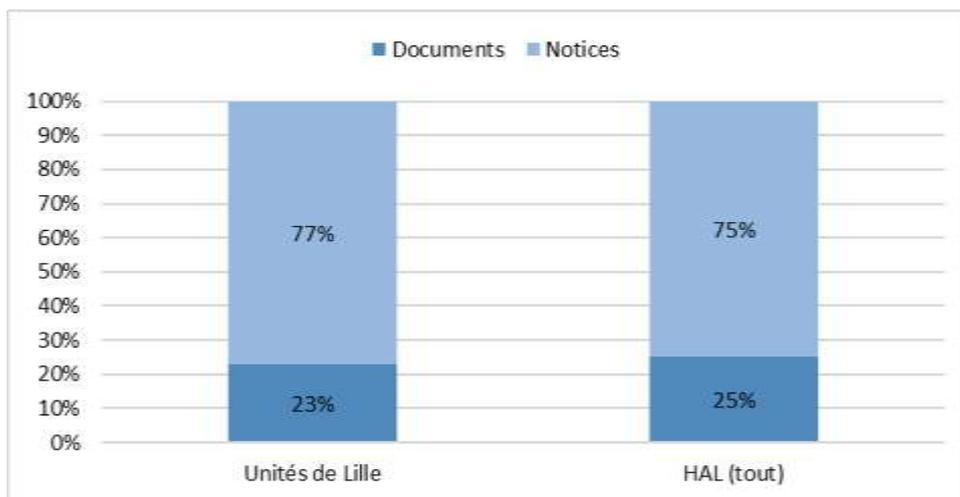


Figure 5. Documents et notices des unités de Lille et de HAL (avril 2020)

Ce taux varie d'un domaine à l'autre. Concrètement, avec beaucoup moins de dépôts que dans les autres domaines, les sciences de la vie et de la santé ont un « taux d'ouverture » bien plus élevé, 52% (figure 6).

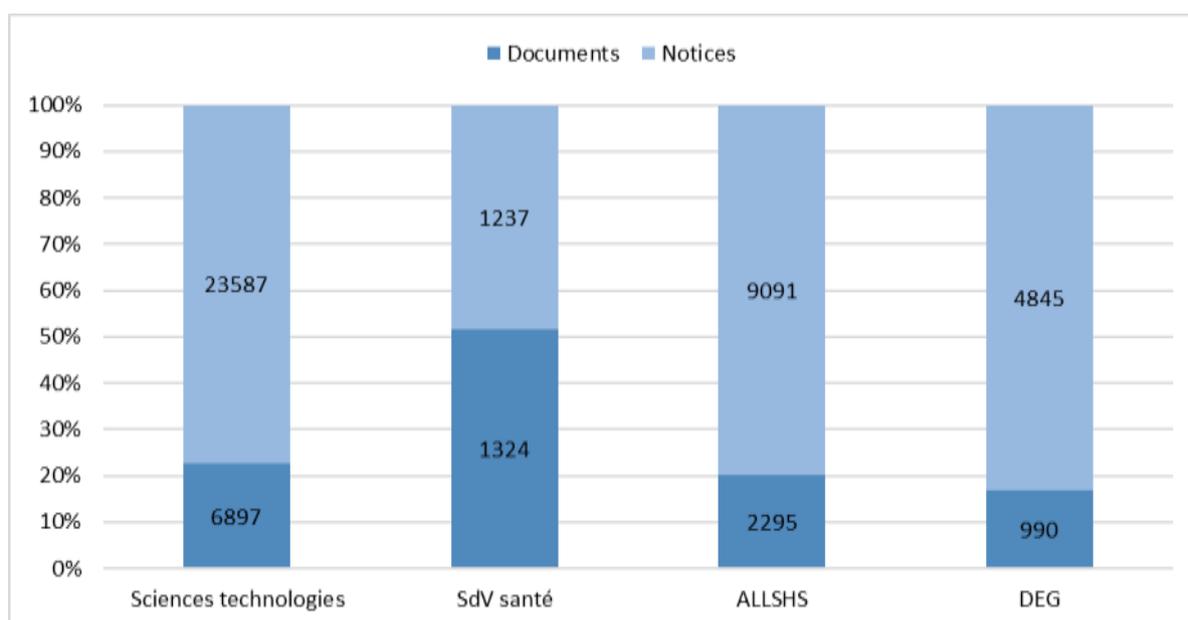


Figure 6. Documents et notices des unités de Lille par domaine (N=50 266)

Quand on regarde de plus près, on se rend compte que là aussi, il s'agit surtout de l'effet d'un petit groupe de cinq unités dont les chercheurs ont déposé plusieurs centaines de documents sur HAL (dont les trois laboratoires *Evaluation des technologies de santé et des pratiques médicales*, *Centre d'infection et d'immunité de Lille* et *Lille Neurosciences & Cognition*).

Également significatif : le taux relativement bas de dépôts avec documents dans les domaines droit, économie et gestion.

La relation entre la création d'une collection et le taux n'est pas évidente, et elle dépend apparemment du domaine (figure 7).

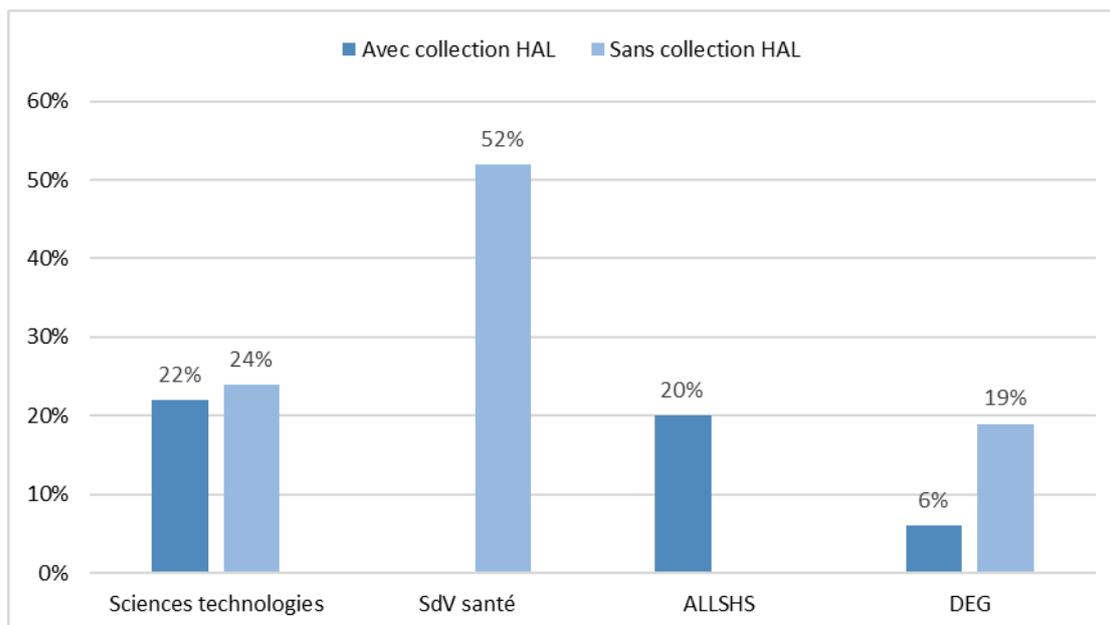


Figure 7. Part des documents par domaine, avec ou sans collection (N=50 266)

Pour les laboratoires en ST, le fait avoir une collection ou non ne change pas le taux d'ouverture. Dans les sciences de la vie et de la santé, il n'y pas de collection ; tandis qu'en SHS, tous les laboratoires ont créé une collection sur HAL, dont le taux global d'ouverture n'est pas très différent de la moyenne générale. D'un point de vue statistique, plusieurs résultats sont significatifs :

- ST : le nombre de dépôts de documents hors collection.
- SdV santé : l'absence d'une collection ; le nombre de dépôts de documents hors collection.
- ALLSHS : le fait que toutes les unités ont créé une collection sur HAL.
- DEG : le nombre de dépôts sans documents (notices), surtout hors collection.

5. Discussion

5.1. Limitations

Quatre aspects limitent potentiellement la portée des résultats de cette étude.

- La dynamique de la situation : L'étude présente des résultats à l'instantanée, pour un moment précis, comme une photographie. Or, la situation des unités de recherche et leur usage de HAL sont dynamiques, avec une alimentation irrégulière, des restructurations et des changements dans l'environnement institutionnel immédiat, dont la création par l'Université de Lille d'un portail HAL¹⁸ et la mise en place d'une archive institutionnelle¹⁹. Une photographie ne peut pas refléter cette dynamique ; mais nous avons essayé de palier à cette limitation par une comparaison avec les résultats de 2019.
- L'absence d'une approche qualitative : Comme en 2019, l'étude se limite à une analyse quantitative des dépôts, sans questionner les chercheurs, documentalistes, ingénieurs ou administratifs et sans analyser les contenus de documents sur le fonctionnement et la production des unités de recherche (rapports d'évaluation, bilans etc.). Ceci limite avant tout l'interprétation des différences entre les laboratoires. La suite de notre projet de recherche intégrera la dimension qualitative (cf. conclusion).

¹⁸ Portail HAL - Université de Lille <https://hal.univ-lille.fr/>

¹⁹ Lille Open Archive LILLOA <https://lilloa.univ-lille.fr/>

- La taille de l'échantillon : Une étude de 62 unités de recherche permet d'identifier et d'illustrer un certain nombre d'approches et de pratiques différentes mais il faut rester prudent par rapport à une généralisation ou transposition des résultats. Pour la même raison, l'étude n'inclut pas d'analyse statistique plus poussée pour le regroupement des unités et pour l'évaluation des relations et effets des différentes variables (analyse factorielle, variance ou régression).
- Le focus sur HAL : L'analyse se limite à l'usage de la plateforme HAL. Or, pour étudier la stratégie des unités de recherche par rapport aux archives ouvertes, il faut tenir compte du fait que leurs chercheurs utilisent aussi d'autres plateformes, comme RePeC²⁰ dans les domaines de l'économie et de la gestion, bioRxiv²¹ et medRxiv²² dans les domaines des sciences de la vie et de la santé, et SocArXiv²³ en sciences sociales. L'enquête auprès des directeurs des laboratoires du CNRS avait déjà révélé que si HAL était bien mentionnée comme la première plateforme d'auto-archivage, elle n'était pas la seule, et que 30% des directeurs évoquaient d'autres plateformes utilisées par les chercheurs de leur laboratoire, dont surtout arXiv²⁴ pour les domaines mathématiques, informatique et physique (Schöpfel et al. 2016). Sur toutes ces plateformes, on trouve des dépôts de plusieurs unités de recherche de l'Université de Lille (dont CIIL, UGSF, Evo-Eco-Paleo et RID-AGE). Mais il reste à mesurer l'importance et la différence de ces dépôts, en comparaison avec ceux de HAL.

5.2. Regroupements

Dans notre étude de 2019, nous avons identifié plusieurs groupes de laboratoires (clusters) qui se distinguent par rapport à l'usage de HAL et notamment, par rapport au nombre de dépôts, à la part des documents en libre accès et à la présence d'une collection. Sans surprise, on retrouve ces clusters aussi dans les résultats de 2020 (figure 8, clusters en rouge).

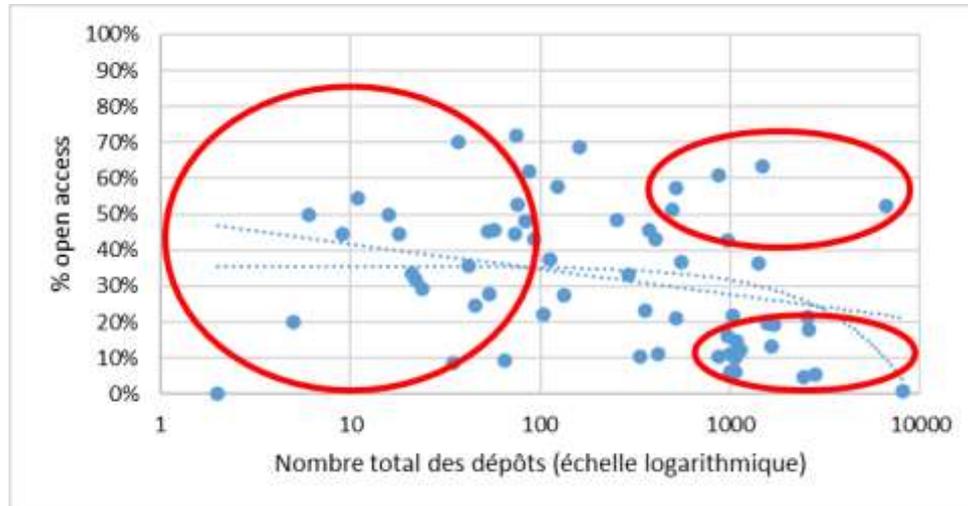


Figure 8. Nombre total des dépôts et % open access (N=62 laboratoires)

- Cluster 1 (39%) : Le premier cluster regroupe les 24 unités de recherche qui totalisent moins de 100 dépôts sur HAL. A part deux unités en DEG et ALLSHS, aucune n'a créé une collection

²⁰ Research Papers in Economics <http://www.repec.org/>

²¹ Preprint Server for Biology <https://www.biorxiv.org/>

²² Preprint Server for Health Sciences <https://www.medrxiv.org/>

²³ Open Archive of the Social Sciences <https://osf.io/preprints/socarxiv>

²⁴ Open archive for scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics <https://arxiv.org/>

sur HAL. D'une certaine manière, on pourrait interpréter ce cluster par l'absence d'une stratégie institutionnelle, au niveau du laboratoire. Les dépôts et la part parfois élevée de documents en libre accès reflètent probablement avant tout l'engagement individuel et l'auto-archivage par les chercheurs eux-mêmes. Quelques exemples : le *Centre d'étude des arts contemporains* (CEAC), le *Centre d'histoire judiciaire* (CHJ), le *Groupe de recherche sur les formes injectables et les technologies associées* (GRITA), le *Laboratoire de mécanique de Lille, multiphysique, multiéchelle* (LaMcube).

- Cluster 2 (24%) : Le deuxième cluster regroupe 15 unités avec un nombre élevé de dépôts (plus de 700) mais peu de documents en libre accès (moins de 20%). Il s'agit donc d'unités dont l'usage de HAL correspond surtout au signalement de leur production scientifique, pour la visibilité et pour l'évaluation. La plupart de ces unités (9) dispose d'une collection sur HAL. Tous les domaines sont représentés, à part les SdV santé. Des exemples : le *Centre d'études en civilisation, langues et lettres étrangères* (CECILLE), le *Centre de recherche droits et perspectives du droit* (CRDP), le *Laboratoire de spectrochimie infrarouge et raman* (LASIR).
- Cluster 3 (11%) : Un petit groupe de 7 unités a un nombre de dépôts supérieur à la médiane (plus de 300 dépôts), avec un pourcentage de documents en libre accès supérieur à la moyenne (plus de 40% dépôts en libre accès). Il s'agit de laboratoires en ST et en SdV santé. Seulement deux unités ont créé une collection. Des exemples : l'*Observatoire de l'Université de Lille, institut de mécanique céleste et de calculs des éphémérides* (IMCCE), le *Laboratoire d'optique atmosphérique* (LOA), le *Centre d'infection et d'immunité de Lille* (LIIL).

L'intérêt de ces regroupements est qu'ils aident à visualiser le caractère multidimensionnel de l'usage de HAL par les unités de recherche, avec des stratégies collectives, des pratiques disciplinaires et individuelles mais aussi des effets structurels, comme le nombre de chercheurs et l'existence d'un service de documentation.

Ces trois clusters regroupent des unités avec un usage similaire de HAL mais ils sont purement descriptifs et n'expliquent rien. Aussi, un quart environ des unités (26%) se trouvent en dehors de ces trois clusters, et il est tout à fait possible d'affiner le regroupement des unités, avec d'autres et davantage de clusters.

5.3. Explications

A partir d'études antérieures et de nos résultats, plusieurs facteurs semblent contribuer à l'usage de HAL par les unités de recherche :

- Le nombre de chercheurs : La taille des unités est fortement corrélée avec le nombre de dépôts. Néanmoins, cette relation est modérée par le domaine scientifique : ainsi, pour les unités en biologie et médecine on pourrait s'attendre, d'après leur taille médiane (20 chercheurs), à un nombre médian de 400 dépôts par unité ; or, en réalité ce chiffre n'est que de 66.
- Un choix délibéré du laboratoire : La création d'une collection de laboratoire sur HAL et l'alimentation de cette collection pour rendre visible la production scientifique du laboratoire sont des indicateurs d'un choix délibéré qu'on peut interpréter comme une stratégie collective, au niveau de l'unité de recherche, d'utiliser la plateforme nationale et ses outils pour contribuer au rayonnement et au développement du laboratoire. Ce choix peut conduire à une action ponctuelle (comme c'est le cas, apparemment, pour l'IEMN) ou à une alimentation dans la durée.
- Des pratiques disciplinaires : Certaines communautés disciplinaires pratiquent davantage la communication scientifique directe via les archives ouvertes que d'autres. Ces traditions et

valeurs en matière de « bonnes pratiques du libre accès » se reflètent partiellement dans des outils, fonctionnalités et partenariats de la plateforme HAL, comme l'alimentation d'arXiv et le moissonnage par RePEc.

- Une politique d'établissement : La politique en faveur de HAL de l'ancienne Université de Lille 3, la création d'un portail d'établissement sur HAL et le soutien des dépôts par l'ancien SCD de Lille 3 ne sont pas étrangers au fait que toutes les unités SHS de l'Université de Lille disposent aujourd'hui d'une collection sur HAL. Au même titre, la politique en faveur de HAL de l'INRIA, le lancement du portail HAL-Inria en 2005, l'obligation de dépôt du texte intégral depuis 2015, la mise à disposition d'outils pour faciliter l'alimentation de HAL etc., tout cela contribue à l'usage significatif de HAL par le principal laboratoire en informatique à Lille, l'unité mixte de recherche CRISTAL.
- Pratiques individuelles : L'engagement individuel des chercheurs en faveur du libre accès et de la plateforme HAL, à titre d'auteur ou en tant que responsable d'un projet, d'une équipe ou d'une unité, joue également un rôle dont l'impact devient visible notamment là où il n'y a pas d'approche institutionnelle ou disciplinaire.

La Figure 9 essaie de visualiser ces variables et leurs liens.

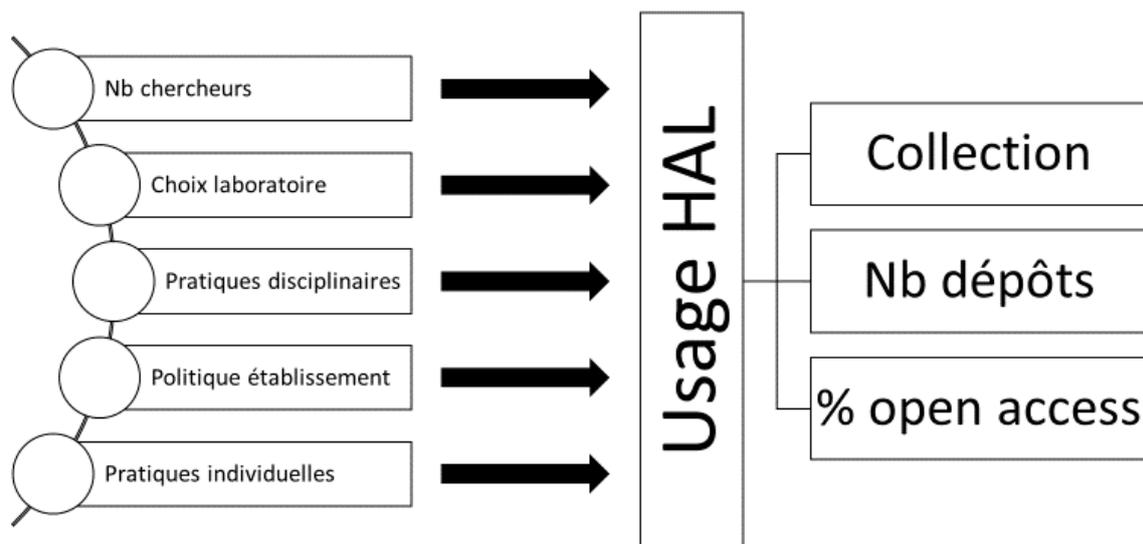


Figure 9. Les variables de l'usage de HAL par les unités de recherche

Nos résultats ne permettent pas de pondérer l'effet de ces différentes variables. Cependant, il paraît probable que ces variables ne sont pas indépendantes au sens strict du terme et de l'analyse statistique. Au contraire, pratiques individuelles, choix collectifs, politiques d'établissements et pratiques disciplinaires font partie d'un écosystème dont les interdépendances et intercorrélations restent à déterminer, aussi par rapport aux ressources mises en œuvre (personnel, budget, équipement et outils...). Il est certain, cependant, que le contexte politique, c'est-à-dire l'engagement du Ministère en faveur de la science ouverte et l'incitation forte des chercheurs à publier en mode « open access », par le biais des revues en libre accès et/ou des archives ouvertes (dont HAL), joue un rôle aussi bien pour les choix institutionnels que pour les décisions individuelles.

5.4. Evolution

Comparée à l'étude de 2019 (juin 2019), l'analyse de 2020 (avril 2020) montre une croissance globale de 21% du nombre de dépôts, moins pour les documents en texte intégral (+20%), plus pour les notices sans documents (+22%). La corrélation entre le nombre de dépôts et leur augmentation est légèrement négative ($r=-.29$), ce qui paraît logique. Ici aussi, on peut identifier plusieurs trajectoires types et quelques cas particuliers (figure 10).

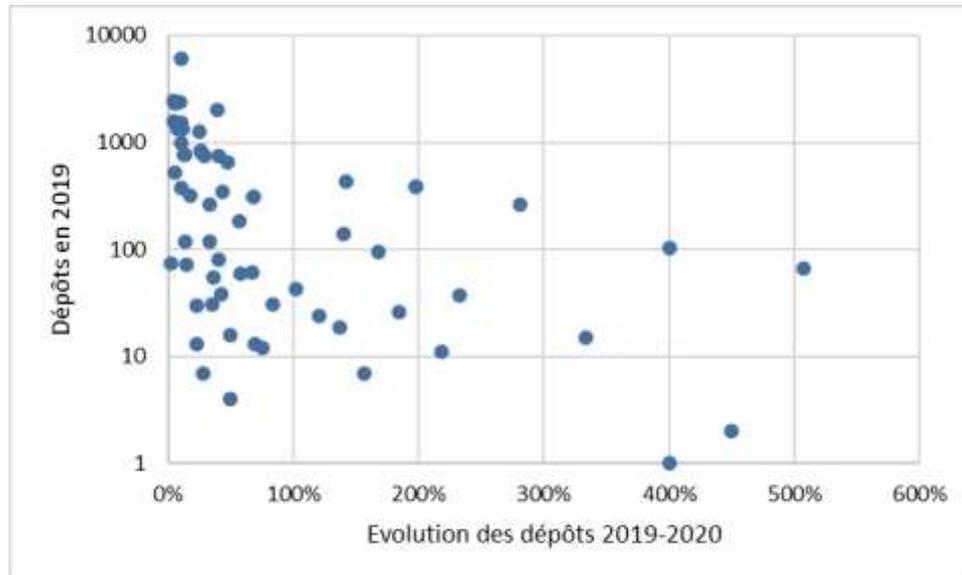


Figure 10. Nombre de dépôts en 2019 (échelle logarithmique) et évolution des dépôts entre 2019 et 2020 (en %)

- D'une part, on peut observer quelques unités avec peu de dépôts sur HAL (<100), dont le nombre augmente très fortement – en pourcentage (jusqu'à +507%).
- D'autres unités, en revanche, avec très peu de dépôts, n'ont pratiquement pas bougé.
- D'autre part, la plupart des unités avec un nombre important de dépôts sur HAL (>300) ont vu une augmentation moyenne ou moins importante, comme CRISTAL (+11%) ou le LASIR (+6%).
- D'une manière plus exceptionnelle, quelques unités avec plus de 100 dépôts ont fortement augmenté le nombre de leurs dépôts sur HAL, comme par exemple le CERAPS (+142%) ou l'UMET (+198%), ce qui correspond probablement à un changement de stratégie par rapport à la plateforme nationale.

Par ailleurs, une seule unité (UCCS) a créé une nouvelle collection sur HAL, en augmentant ses dépôts de 40%. Malgré ces trajectoires et approches différents, l'effet sur le pourcentage des documents en libre accès est minime. Même si le nombre de ces documents a augmenté depuis 2019 de 2 052 (+22%), leur part relative à l'ensemble des dépôts stagne : il était de 22,7% en juin 2019, il est de 22,9% en avril 2020. Le nombre des notices (sans documents), quant à elles, a augmenté en même temps de 6 513.

5.5. L'usage dans le contexte

Notre analyse porte sur l'usage de la plateforme HAL par les unités de recherche. L'objectif est d'apporter quelques éléments empiriques à la connaissance de la « conduite face à l'objet » HAL, l'emploi de cet outil (Jouët 1993) par les laboratoires comme des structures intermédiaires entre les établissements et organismes et les chercheurs à titre individuel.

A ce stade, nous n'avons pas voulu observer les pratiques des chercheurs, par une approche ethnographique de la simple utilisation, c'est-à-dire, de la « manière de faire » de la plateforme, ou par une analyse des déclarations des acteurs eux-mêmes. L'originalité de notre approche réside dans l'étude des usages de HAL à partir de l'analyse des résultats, opérationnalisés sous forme de plusieurs variables dépendantes (dépôts, notices, création de collection...).

A notre avis, ce sont ces résultats qui donnent du sens à l'usage de HAL, compris comme un continuum entre l'appropriation d'un outil encore relativement récent et nouveau par certains, et son utilisation plus confirmée (au sens d'une pratique sociale) par d'autres, un procès à la fois individuel et social (Proulx 2005, Paganelli 2012).

Il s'agit d'une analyse préliminaire, dans le cadre d'un projet plus large qui, lui, inclura la dimension sociale de l'usage (cf. Gardiès et al. 2010), par l'étude des pratiques, attitudes, représentations et compétences des acteurs dans les laboratoires - chercheurs, administratifs, ingénieurs, documentalistes... (Proulx 2005). La réanalyse de l'enquête auprès des directeurs des laboratoires du CNRS avait partiellement déjà abordé cet aspect (Schöpfel et al. 2016), faisant le lien entre les pratiques déclarées, les attitudes (opinions) et l'évaluation des compétences. Dans le cadre du projet en cours, il s'agira de faire le rapprochement entre les résultats visibles sur HAL et la représentation, les explications et les attitudes des acteurs impliqués, par le biais d'enquêtes et d'entretiens. Une telle approche devra tenir compte du « jeu réflexif » du fait que les chercheurs observateurs font partie de la situation, partagent (certaines) pratiques et (certaines) attitudes (Proulx 2015). Ce « jeu réflexif » est d'ailleurs l'une des raisons pour avoir commencé l'étude des usages de HAL par les unités de recherche par une analyse de notre propre laboratoire (Schöpfel et al. 2018).

6. Conclusion

Le point de départ de notre étude était la question de l'usage que font les unités de recherche de la plateforme HAL. A partir de la présence sur HAL des 62 laboratoires de l'Université de Lille, l'analyse révèle une large diversité d'approches et de pratiques, en particulier par rapport à la mise en place d'une collection sur HAL, par rapport au nombre des dépôts et par rapport à la part des documents diffusés en texte intégral sur HAL. L'analyse montre des différences disciplinaires et des similitudes d'approches entre différents laboratoires. Elle contribue ainsi à une meilleure connaissance et compréhension des pratiques en matière du libre accès en général et de l'appropriation de la plateforme HAL par les unités de recherche en particulier.

A la différence d'autres études, notre recherche met l'accent sur l'unité de recherche (le laboratoire) et non pas sur le chercheur en tant qu'individu ou sur l'établissement (université, organisme de recherche) comme tutelle. L'intérêt d'une telle approche conceptuelle et méthodologique est la vue d'ensemble, la dimension systémique et fonctionnel où les politiques institutionnelles, les pratiques disciplinaires et les choix et engagements individuels s'entrecroisent pour former (ou non) une stratégie structurante au niveau de l'unité de recherche.

Pour la suite, nous menons actuellement un projet de recherche avec les plus de 1 500 laboratoires des universités du programme IDEX²⁵, avec pour chaque unité une étude détaillée des différents types et modalités de dépôts et en incluant les dimensions qualitatives (enquêtes, entretiens) et institutionnelles (statuts et tutelles des unités) pour mieux comprendre les choix et leurs déterminants. Les résultats sont attendus pour fin 2021.

²⁵ Projet HAL/LO – Valorisation sur HAL de la production des laboratoires dans l'environnement de la science ouverte http://gis-reseau-urfist.fr/?page_id=154

Pour pousser plus loin l'analyse de l'écosystème de la science ouverte, on pourrait aussi étendre l'analyse des stratégies institutionnelles à d'autres structures et organisations intermédiaires, dont notamment les groupements et fédérations de recherche et, dans la ligne d'un rapport récent (Egret & Fabre 2019), les infrastructures de recherche. Quel est leur rôle pour le développement de la science ouverte en France ? Comment contribuent-ils à la communication scientifique directe et à l'appropriation de la plateforme HAL par les équipes et individus ? Comment mesurer leur impact dans le millefeuille de l'ESR français ? Il serait particulièrement intéressant, dans ce contexte, d'évaluer l'effet de plusieurs événements et développements significatifs sur ces stratégies et approches, dont le lancement du portail HAL du CNRS²⁶, le projet de refondation de HAL et, à un tout autre niveau, la pandémie de la Covid-19.

Remerciements

L'étude a été menée dans le cadre du projet HAL/LO, avec un financement par le GIS URFIST 2019-2021. Mes remerciements aux membres de l'équipe de projet pour leurs questions et conseils utiles.

Références

- ARLITSCH K., GRANT C., « Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape ». *Journal of Library Administration*, 2018, vol. 58, n° 3, p. 264–281. <https://doi.org/10.1080/01930826.2018.1436778>
- BARUCH P., « Open access developments in France: the HAL Open Archives System ». *Learned Publishing*, 2007, vol. 20, n° 4, p. 267–282. <https://doi.org/doi:10.1087/095315107x239636>
- CHARNAY D., MICHAU C. L'archive ouverte HAL. In : JRES 2007, Strasbourg, Novembre 2007. <https://hal.archives-ouvertes.fr/hal-01101888>
- CHARTRON G., SCHÖPFEL J. « Open access et Open science en débat ». *Revue Française Des Sciences de l'information et de La Communication*, 2017, n° 11. <https://doi.org/10.4000/rfsic.3331>
- CNRS. "Mieux partager l'IST" - Enquête nationale conjointe auprès des Universités, Organismes et Ecoles d'Ingénieurs sur le Partage et la Gestion des ressources d'IST. Rapport. Paris : CNRS Direction IST, 2016. <https://www.science-ouverte.cnrs.fr/dist-dossiers/>
- CNRS. L'articulation des archives des établissements et de l'archive nationale pluridisciplinaire HAL. Etude COPIST 4. Rapport. Paris : CNRS Direction IST, 2018. <http://adbu.fr/les-etudes-du-copist-catalogue-doffres-partagees-en-ist/>
- COAR. *Promoting Open Knowledge and Open Science. Report of the Current State of Repositories*. Göttingen : Confederation of Open Access Repositories, 2015. <https://www.coar-repositories.org/files/Promoting-Open-Knowledge-and-Open-Science-Report-on-the-current-state-of-Repositories.pdf>
- COUPERIN. *Enquête sur les archives ouvertes françaises menée par Couperin en 2019. Résultats et analyses*. Rapports et études Couperin N° 2. Paris : Consortium Couperin, 2020. <https://www.couperin.org/site-content/261-a-la-une/1407-resultats-de-l-enquete-sur-les-pratiques-de-publication-et-d-acces-ouvert-des-chercheurs-francais>
- EGRET D., FABRE R. *Mission "Publimétries". Apprécier l'impact scientifique des TGIR/OI "Comment mieux tracer les dynamiques du travail scientifique."* Paris : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2019. <http://perso.obspm.fr/daniel.egret/Publimetriques.pdf>

²⁶ HAL-CNRS Publications scientifiques du CNRS <https://hal-cnrs.archives-ouvertes.fr/>

- GANDOSSI A. *Usages, pratiques et besoins des chercheurs concernant les serveurs d'archives ouvertes*. Mémoire de Master. Université de Lille 3, 2011. https://memsic.ccsd.cnrs.fr/mem_00655746
- GARDIES C., FABRE I., COUZINET V., « Re-questionner les pratiques informationnelles ». *Études de Communication*, 2010, n° 35, p. 121–132. <https://doi.org/10.4000/edc.2241>
- GAYOSO E. *La diffusion sur Hal, Academia et ResearchGate des articles de recherche des revues françaises de Sciences Humaines et Sociales*. Rapport. Paris : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2020. <https://www.enseignementsup-recherche.gouv.fr/cid136723/le-soutien-a-l-edition-scientifique.html>
- HARNAD S. et al., « The Access/Impact Problem and the Green and Gold Roads to Open Access ». *Serials Review*, 2004, vol. 30, n° 4, p. 310–314. <https://doi.org/10.1016/j.serrev.2004.09.013>
- JOUËT, J. « Usages et pratiques des nouveaux outils de communication ». In : SFEZ L. (dir.), *Dictionnaire critique de la communication* vol. 1 (pp. 371–376). Paris : PUF, 1993. P. 371-376.
- LAFON S. *Les Archives Ouvertes en France et le rôle des professionnels de l'information de l'Enseignement supérieur et de la Recherche*. Mémoire de Master. Université Paul Valéry Montpellier 3, 2013(a). http://www.sylvie-lafon.fr/var/f/1s/sw/1sswN5J9PWKMzje-Y63BlknSwTphiyVLIQrmX8UfxCa_master.pdf
- LAFON S. *Le développement de l'archive ouverte institutionnelle HAL-UPS : Préconisations pour la mise en place d'un workflow pour la chaîne de traitement documentaire des publications scientifiques des laboratoires de recherche de l'Université Toulouse III – Paul Sabatier*. Rapport. Université de Toulouse 3, 2013(b). https://hal.archives-ouvertes.fr/sic_00977548
- LARROCHE V. *Le dispositif : un concept pour les sciences de l'information et de la communication*. London : ISTE Editions, 2018.
- MAGRON A. « Utiliser les archives ouvertes pour valoriser ses travaux l'exemple de HAL-SHS ». In : SCHNEDECKER C., ALEKSANDROVA A. (dir.). *Le doctorat en France : mode(s) d'emploi*. Berlin : Peter Lang, 2017. p. 227–239. https://hal.archives-ouvertes.fr/sic_01697164
- MAHÉ A. « Les pratiques informationnelles des chercheurs dans l'enseignement supérieur et la recherche : regards sur la décennie 2000-2010 ». In : CHARTRON G., EPRON B., MAHE A. (dir.). *Pratiques documentaires numériques à l'université*. Villeurbanne : Presses de l'Esssib, 2017. p. 12–41. <https://doi.org/10.4000/books.pressesenssib.1171>
- MAHÉ A., PRIME-CLAVERIE C., « 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*, 2017(a), vol. 11. <https://doi.org/10.4000/rfsic.3315>
- MAHÉ A., PRIME-CLAVERIE C., « Science ouverte et présence numérique des chercheurs en sciences humaines et sociales Une étude exploratoire à partir de deux plateformes en ligne : HAL-SHS et Hypotheses.org ». *Document Numérique*, 2017(b), vol. 20, n° 2–3, p. 79–96. <https://doi.org/10.3166/dn.2017.00010>
- MESRI. *Plan national pour la science ouverte*. Paris : Ministère de l'Enseignement supérieur, de la recherche et de l'innovation, 2018. <https://www.ouvrirelascience.fr/plan-national-pour-la-science-ouverte/>

PAGANELLI C. « Analyse des discours sur la notion d'usage' dans deux revues en sciences de l'information : Doc-SI et BBF ». *Documentaliste-Sciences de l'Information*, 2012, vol. 49, n° 2, p. 64-71. <https://doi.org/10.3917/docsi.492.0064>

PRIME-CLAVERIE C., MAHÉ A. « Sites de dépôt en libre accès et formes de médiations : quelles évolutions ? » In : BOUSTANY J., BROUDOUX É., CHARTRON G. (dir.). *La médiation numérique : renouvellement et diversification des pratiques. Actes du colloque Document numérique et société, Zagreb 2013*. Bruxelles : De Boeck, ADBS, 2013. p. 125–139. <https://doi.org/10.3917/dbu.chron.2013.01.0125>

PROULX S. « Penser les usages des TIC aujourd'hui : enjeux, modèles, tendances ». In : VIEIRA L., PINEDE N. (dir.), *Enjeux et usages des TIC : aspects sociaux et culturels t.1*. Bordeaux : Presses universitaires de Bordeaux, 2005. p. 7-20.

PROULX S. « La sociologie des usages, et après ? » *Revue Française Des Sciences de l'information et de La Communication*, 2015, (6). <https://doi.org/10.4000/rfsic.1230>

ROBINSON-GARCIA N., COSTAS R., VAN LEEUVEN T. N., « Open Access uptake by universities worldwide ». *PeerJ*, 2020. <https://doi.org/10.5281/zenodo.3874959>

SCHÖPFEL J., FERRANT C., ANDRÉ F., FABRE R., « Ready for the Future? A Survey on Open Access with Scientists from the French National Research Center (CNRS) ». *Interlending & Document Supply*, 2016, vol. 44, n° 4, p. 141–149. <https://doi.org/doi:10.1108/ILDS-06-2016-0023>

SCHÖPFEL J., PROST H., FRAISSE A., CHAUDIRON S. Valoriser les publications d'un laboratoire universitaire dans l'environnement de la science ouverte : Retour d'expérience de la collection GERiiCO sur HAL. In : ICOA 2018 3e Colloque International sur le Libre Accès, Rabat, Maroc, 28-30 Novembre 2018. <https://hal.archives-ouvertes.fr/hal-01940352>

SCHÖPFEL J., PROST H., NDIAYE I. E. H. Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL. In : GL 21 International Conference on Grey Literature, Hannover, Allemagne, 22-23 octobre 2019. <https://hal.archives-ouvertes.fr/hal-02300017>

SMITH C. *Formation au dépôt dans HAL. Archive Ouverte HAL-LAAS*. Support de formation. Toulouse : CNRS, 2017. <https://hal.laas.fr/cel-01670071>

WOJCIECHOWSKA A. « Pratiques documentaires et pratiques d'auto-archivage des mathématiciens et informaticiens en France ». In CHARTRON G., EPRON B., MAHÉ A. (dir.). *Pratiques documentaires numériques à l'université*. Villeurbanne : Presses de l'Ensib, 2017. p. 90–120. <https://books.openedition.org/pressesenssib/1176>