

## **Balisages**

ISSN : 2724-7430

Éditeur : Ecole nationale supérieure de l'information et des bibliothèques

6 | 2023

30 ans après le Web... Big bang dans les institutions de savoirs?

---

# L'utilisation de HAL par les laboratoires de recherche

Une étude quantitative

**Joachim Schöpfel, Florence Thiault, Hélène Prost, Bernard Jacquemin et  
Éric Kergosien**

---

🔗 <https://publications-prairial.fr/balisages/index.php?id=1166>

DOI : 10.35562/balisages.1166

### **Référence électronique**

Joachim Schöpfel, Florence Thiault, Hélène Prost, Bernard Jacquemin et Éric Kergosien, « L'utilisation de HAL par les laboratoires de recherche », *Balisages* [En ligne], 6 | 2023, mis en ligne le 21 septembre 2023, consulté le 06 octobre 2023.  
URL : <https://publications-prairial.fr/balisages/index.php?id=1166>

### **Droits d'auteur**

CC BY SA 4.0

# L'utilisation de HAL par les laboratoires de recherche

Une étude quantitative

Joachim Schöpfel, Florence Thiault, Hélène Prost, Bernard Jacquemin et  
Éric Kergosien

## PLAN

---

Introduction

État de l'art

Archives ouvertes et archives institutionnelles

Pratiques individuelles et pratiques collectives

Le rôle des laboratoires

Un paysage hétérogène

L'impact des disciplines

Des facteurs interdépendants

Méthodologie

Résultats

L'utilisation de HAL

Typologie des dépôts

Documents en texte intégral

Collections

Discussion

Le rôle des laboratoires

L'impact des disciplines scientifiques

L'évolution du dispositif HAL

Conclusion

Annexes

Le statut des 1 246 laboratoires

L'affiliation universitaire des laboratoires

Les domaines des laboratoires

Les disciplines des laboratoires (RNSR)

## TEXTE

---

*Remerciements : Le projet HAL/LO a reçu un financement dans le cadre d'un appel à projet du Groupe d'intérêt scientifique (GIS) « Réseau Urfist ». Nous remercions également les deux peer reviewers anonymes de la revue Balisages pour leurs commentaires et recommandations attentives.*

# Introduction

- 1 La France a inscrit la science ouverte dans les missions des chercheurs et des enseignants-chercheurs et poursuit en particulier l'objectif de généraliser l'obligation de publication en accès ouvert des articles et livres issus de recherches financées par appel à projets sur fonds publics<sup>1</sup>, que ce soit par une publication nativement en accès ouvert ou par le dépôt dans une archive ouverte. L'archive ouverte pluridisciplinaire HAL se trouve ainsi confortée au cœur de la politique en faveur de la science ouverte en France (Berthaud *et al.*, 2021).
- 2 Quel est l'intérêt de HAL pour les laboratoires de recherche comme pivot intermédiaire et structurant de la recherche scientifique en France, porteurs de projets de recherche et objets de l'évaluation ? Comment font-ils pour gérer et valoriser leur production scientifique sur HAL ? Initialement, les archives ouvertes ont été créées pour l'auto-archivage des publications par leurs auteurs (Harnad *et al.*, 2004). Aujourd'hui, ces mêmes archives sont aussi alimentées par d'autres personnels (documentalistes, bibliothécaires, etc.), par d'autres plateformes et bases de données, et par les éditeurs eux-mêmes (Charnay, 2019). La gestion se fait à plusieurs niveaux, avec des collections, des portails ou encore des pages personnelles. Stratégies institutionnelles et pratiques personnelles cohabitent, dans un écosystème disciplinaire et organisationnel de plus en plus complexe.
- 3 Afin de mieux comprendre le rôle des laboratoires dans cet environnement, nous avons conduit un projet de recherche sur les stratégies de valorisation de la production scientifique des laboratoires de recherche sur HAL<sup>2</sup>. À ce jour, nous avons publié une analyse de la typologie des dépôts (Schöpfel *et al.*, 2020 ; 2021), les résultats d'une enquête sur l'appropriation de HAL par les laboratoires (Schöpfel *et al.*, 2022) et une analyse scientométrique de l'alimentation de HAL par les laboratoires (Schöpfel *et al.*, 2023).
- 4 Dans le cadre de ce projet, nous avons mené plusieurs analyses quantitatives en 2021 et 2022 pour répondre à quatre questions :
  1. Est-ce que tous les laboratoires utilisent HAL ?
  2. Combien de laboratoires déposent quels types de documents ?
  3. Combien de documents sont déposés en texte intégral ?

4. Combien de laboratoires ont créé une collection sur HAL ?
- 5 L'objectif de ces analyses était de produire quelques résultats empiriques pour décrire d'une manière plus détaillée comment les laboratoires de recherche utilisent HAL ; ces analyses viennent compléter d'autres études en proposant plus particulièrement des approches qualitatives.
- 6 La discussion des résultats portera sur le rôle des laboratoires dans l'environnement de la science ouverte, sur les différences disciplinaires et sur la transformation de HAL d'une archive ouverte vers un dispositif d'évaluation.

## État de l'art

### Archives ouvertes et archives institutionnelles

- 7 Les archives ouvertes ont été décrites comme des bases de données documentaires accessibles librement et gratuitement sur Internet, dans lesquelles « les chercheurs peuvent y déposer un document scientifique ou technique dont ils sont l'auteur et dont ils détiennent les droits de diffusion pour rendre ce document librement accessible », de manière à ce que « les lecteurs (puissent) rechercher, consulter et télécharger librement et gratuitement un document scientifique qui (y) a été déposé » (Deboin, 2015). Il s'agit de plateformes qui fonctionnent sur le mode de l'auto-archivage, c'est-à-dire, « le téléchargement, ou “dépôt”, d'un document par son auteur sur le site web de l'archive ouverte » (Deboin, 2015 ; Harnad *et al.*, 2004). Les premiers sites de ce type ont été créés il y a 30 ans ; aujourd'hui (août 2022), le répertoire international OpenDOAR<sup>3</sup> recense plus de 6 000 archives ouvertes, dont 162 en France.
- 8 Neuf sur dix de ces sites correspondent à des archives institutionnelles, définies comme des services qu'une institution (université, organisme scientifique, etc.) « offre aux membres de sa communauté pour la gestion et la diffusion de ressources numériques créées par l'institution et les membres de sa communauté » (Lynch, 2003), avec des avantages aussi bien pour l'institution (archivage, mise en valeur

de la qualité du capital intellectuel, promotion, visibilité et prestige, etc.) que pour les chercheurs (préservation des publications, impact, bilans/CV, etc.) (Westell, 2003 ; Yeates, 2003).

- 9 Créé en 2001 à l'instar du serveur de preprint d'arXiv comme une archive ouverte d'auto-archivage, HAL se positionne aujourd'hui comme une « infrastructure publique, pérenne et responsable » pour la diffusion en libre accès et la préservation à long terme des publications issues de la recherche française<sup>4</sup>, comme un site agrégateur et multidisciplinaire<sup>5</sup>, une « véritable vitrine de la production française »<sup>6</sup> avec un fort impact et un rayonnement international, et avec des fonctionnalités et services à destination des chercheurs et des structures scientifiques.

## Pratiques individuelles et pratiques collectives

- 10 La plupart des études sur l'usage des archives ouvertes en général et de HAL en particulier mettent le focus sur les pratiques individuelles. Ainsi, en sciences humaines et sociales, une étude a identifié trois groupes de chercheurs : ceux qui déposent ponctuellement sur HAL (48 %), ceux qui déposent régulièrement (35 %) et ceux qui déposent de temps en temps (9 %), (Mahé & Prime-Claverie, 2017b).
- 11 Les auteurs ont la possibilité de déterminer les droits de réutilisation lors du dépôt de leurs publications sur HAL. Tandis que le ministère et la plupart des organismes et établissements incitent à partager les documents et données sans restriction, dans le domaine public ou avec une licence libérale (CC-BY, Licence Ouverte Etalab), neuf sur dix dépôts en texte intégral restent protégés par le droit d'auteur ou le copyright (88 %) ; seulement 11 % sont diffusés avec une licence Creative Commons (dont 5,5 % avec CC-BY) ou Etalab, et seulement 1 % sont transférés dans le domaine public (Schöpfel *et al.*, 2021), un choix individuel en contradiction (ou en ignorance) avec les stratégies politiques et institutionnelles en faveur de la science ouverte.
- 12 Cependant, les pratiques individuelles n'expliquent qu'une partie des dépôts sur HAL, celle qui correspond à l'auto-archivage des publications par les auteurs eux-mêmes (la « voie verte », voir Harnad *et al.*, 2004). Or, une part des contenus sur HAL ne correspond pas au prin-

cipe de l'auto-archivage mais à un dépôt par d'autres intermédiaires ou moyens, tout comme sur d'autres archives ouvertes (*mediated archives*, voir Neugebauer & Murray, 2013). En sciences de la vie, moins de la moitié des publications sur HAL est déposée par les auteurs eux-mêmes ; le reste est déposé par d'autres personnels des laboratoires, bibliothèques ou services communs de la documentation (SCD), notamment pour les bilans des laboratoires ou dans le cadre des campagnes d'évaluation (Prime-Claverie & Mahé, 2013). Une étude récente a identifié six types de comptes-contributeurs pour alimenter HAL, en plus des auteurs eux-mêmes ; ces dépôts (qu'on pourrait qualifier d'« allo-archivage », en opposition à l'auto-archivage) correspondent à 48 % du contenu de HAL et témoignent d'une utilisation collective et institutionnelle de HAL à plusieurs niveaux, par des équipes scientifiques, des laboratoires, des SCD, etc., et sous forme de création de notices, d'import de listes de publications, de flux de migration ou d'intégration de métadonnées en provenance d'éditeurs (Schöpfel *et al.*, 2023). On est loin d'une pratique individuelle ; et on est témoin ici d'une évolution de HAL d'un serveur d'auto-archivage vers un dispositif de suivi et d'évaluation de la recherche.

- 13 Une autre étude confirme le lien étroit entre les services de HAL et les pratiques des chercheurs et autres contributeurs : à partir du moment où l'opérateur de HAL, le CCSD, a facilité et simplifié le dépôt, répondant ainsi à une demande forte de la part des communautés scientifiques et des autorités, la qualité des métadonnées baisse, ce qui appauvrit la description des publications sur HAL et rend le référencement et la recherche des contenus plus difficile et aléatoire (Tabariès, 2022).

## **Le rôle des laboratoires**

- 14 Les résultats d'une enquête de l'université de Toulouse montrent que les laboratoires jouent tout d'abord un rôle de communication et de médiation, pour informer et promouvoir la plateforme HAL et pour inciter les chercheurs à y déposer leurs publications (Lafon, 2013). Cette action est complémentaire avec celle menée par les SCD et les directions Recherche (Couperin, 2020), par rapport aux orientations et choix institutionnels des tutelles, notamment – pour les unités

mixtes – du Centre national de la recherche scientifique (CNRS), de l'Institut national de la recherche agronomique (INRAE) et de l'Institut national de recherche en sciences et technologies du numérique (INRIA) (Schöpfel *et al.*, 2022).

- 15 Selon une enquête du CNRS, les directeurs de laboratoires soutiennent majoritairement la plateforme HAL (Schöpfel *et al.*, 2016). Ce soutien institutionnel peut se concrétiser par la création d'une collection HAL et/ou par le signalement systématique de la production scientifique d'un laboratoire sur HAL (Lafon, 2013). L'existence d'une collection témoigne toujours d'un choix significatif, collectif et institutionnel au sein du laboratoire ; elle fait suite à une décision en faveur de l'utilisation de HAL pour afficher la production scientifique, une décision prise par la direction, le conseil ou bureau, et/ou en assemblée générale.
- 16 Mais l'enquête du CNRS a également montré que tous les laboratoires ne sont pas dans la même situation face aux enjeux de la science ouverte – il y a des différences disciplinaires, certes, mais il y a aussi des inégalités en termes d'équipement, de ressources matérielles et humaines (Schöpfel *et al.*, 2016 et 2018b).

## Un paysage hétérogène

- 17 Si la corrélation entre le nombre de chercheurs et le nombre de dépôts sur HAL est forte, ce rapport est modéré par le domaine scientifique et il existe des différences significatives entre les laboratoires qui ne s'expliquent pas par la taille de l'unité (Schöpfel, 2020) :
- certains laboratoires utilisent HAL comme une vitrine pour leur production scientifique, avec une collection qui regroupe beaucoup de publications et un taux élevé de documents en accès libre ;
  - d'autres utilisent HAL comme un gestionnaire de références bibliographiques, avec un nombre important de notices mais peu de documents en texte intégral ;
  - un troisième groupe ne semble pas avoir de stratégie par rapport à HAL, dans la mesure où il n'y a pas de collection et très peu de dépôts.
- 18 Ces *clusters* évoluent différemment : le nombre des laboratoires avec une stratégie de vitrine augmente tandis que la taille des deux autres *clusters* diminue, et de plus en plus de laboratoires se trouvent

quelque part « au milieu » du tableau, partagés entre une approche « vitrine » et une approche « gestionnaire », ce qui peut expliquer (surtout pour les unités mixtes) la décision du CNRS d'obliger les chercheurs de produire leur bilan individuel à partir de HAL. Mais les trajectoires des laboratoires par rapport à HAL sont assez divergentes et en partie aussi liées à des conditions locales (Kergosien & Schöpfel, 2021).

- 19 Une enquête qualitative auprès de 144 laboratoires confirme l'importance des conditions particulières, mais relativise conjointement l'impact des tutelles : « L'impact de cet environnement sur les choix des laboratoires est significatif mais ne détermine pas tout. Par exemple, dans un environnement perçu comme favorable, il y a davantage de laboratoires avec une stratégie définie et davantage de laboratoires avec une collection sur HAL qu'ailleurs. Mais même dans un tel environnement, certains laboratoires n'ont ni stratégie, ni collection » (Schöpfel *et al.*, 2022). Certains répondants semblent tout simplement ignorer les politiques et les services des tutelles en matière de science ouverte, un « "gap" entre une stratégie politique et la visibilité (ou perception) sur le terrain (qui) n'est (pas) une particularité française » (*idem* ; Morais *et al.*, 2021). Une particularité française réside en revanche dans la diversité des tutelles, notamment les différences entre les stratégies du CNRS et des universités, qui, d'après les répondants, n'aident pas les laboratoires à déterminer leurs propres pratiques.
- 20 D'après la même enquête et en regroupant les réponses, on peut différencier les laboratoires d'après trois groupes de variables (Schöpfel *et al.*, 2022) :
- communauté : ce premier groupe de variables correspond aux pratiques des différentes communautés scientifiques (disciplines, thématiques, équipements, etc.) sans lien direct avec les choix d'un laboratoire particulier : la connaissance de HAL et les pratiques des chercheurs (dépôts HAL, utilisation de HAL pour les pages personnelles) ;
  - engagement : le deuxième groupe de variables correspond aux choix et actions mises en œuvre par un laboratoire : la création d'une collection HAL, la définition d'une stratégie, l'information sur HAL, l'organisation de formations ;



- valorisation : le troisième groupe de variables correspond à l'emploi de HAL pour les besoins du laboratoire : l'utilisation de HAL pour les bilans, pour les évaluations Hcéres et/ou sur le site du laboratoire.
- 21 À chaque niveau, on peut identifier plusieurs *clusters* de laboratoires, plus ou moins engagés par rapport à HAL, valorisant HAL peu ou prou pour les besoins de l'unité, regroupant une communauté de chercheurs plus ou moins partie prenante par rapport à HAL. Ces trois niveaux sont interconnectés. 20 % des laboratoires par exemple sont plus engagés et plus valorisants que les autres, sans qu'il y ait un rapport avec la pratique des chercheurs eux-mêmes.

## L'impact des disciplines

- 22 Dès le début de HAL, plusieurs études ont constaté des différences disciplinaires des dépôts, notamment par rapport à la physique et aux mathématiques (Baruch, 2007). On y retrouve « de grandes variations disciplinaires, les domaines les plus propices étant la physique, les mathématiques et l'informatique » (Mahé, 2017). Les directeurs d'unités du CNRS en mathématiques, physique et informatique soutiennent davantage HAL que dans d'autres disciplines (Schöpfel *et al.*, 2016).
- 23 Ces différences impactent d'une manière significative non seulement le nombre des dépôts, mais aussi la distribution des types de publications, la langue des publications et le choix des éditeurs ; comparé au Web of Science, HAL contient moins de médecine et davantage de SHS (Larrieu & Schöpfel, 2022). De même, la corrélation entre le nombre de chercheurs et le nombre de dépôts sur HAL est modérée par le domaine scientifique ; « 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 » (Schöpfel, 2020).
- 24 Ce constat rejoint l'observation par Mahé et Prime-Claverie (2017) 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 ». Il est également compatible avec l'étude sur les comptes-

contributeurs qui révèle que l'auto-archivage est plus important en SHS que dans d'autres domaines (Schöpfel *et al.*, 2023).

- 25 D'une manière plus globale, ces différences sont visibles aussi dans les trois niveaux d'engagement, de valorisation et de pratiques communautaires mentionnés plus haut, mais sans signification statistique (Schöpfel *et al.*, 2022). Par ailleurs, on peut se demander si la politique du CNRS et d'autres organismes ne contribuera pas à terme à estomper une partie de ces différences, en obligeant tous les chercheurs aux mêmes pratiques.

## Des facteurs interdépendants

- 26 Basés sur ces études et analyses, nous avons proposé cinq variables pour expliquer les différences constatées par rapport à l'existence d'une collection, au nombre des dépôts et à la part des documents accessibles (Schöpfel, 2020) :

- le nombre de chercheurs ;
- un choix délibéré du laboratoire ;
- des pratiques disciplinaires ;
- une politique d'établissement ;
- des pratiques individuelles.

- 27 Ces variables ne sont pas indépendantes ; les chercheurs avec leurs pratiques individuelles, les équipes et laboratoires avec leurs choix collectifs, les établissements et les organismes avec leurs politiques, et les communautés disciplinaires avec leurs valeurs et habitudes, tous font partie du même environnement de la recherche publique. Le rôle de HAL varie d'un laboratoire à l'autre, « en fonction des ressources et personnels, des engagements et initiatives, mais aussi de l'environnement institutionnel et des tutelles et des "traditions" propres aux unités et aux communautés [...] Les uns y voient une opportunité pour accroître la visibilité des travaux des chercheurs, les autres considèrent HAL comme une contrainte de plus » (Schöpfel *et al.*, 2021b).

- 28 Malgré la diversité des études et analyses, nous pouvons résumer l'état de l'art en six constats :

1. lancé comme une archive ouverte à l'instar d'arXiv (voie verte), HAL se positionne aujourd'hui comme une sorte d'archive institutionnelle à l'échelle nationale (vitrine de la production française) avec des fonctionnalités et services à destination des chercheurs et des structures scientifiques ;
  2. l'auto-archivage (le dépôt par les chercheurs eux-mêmes) n'explique qu'une partie des contenus de HAL. La part importante de dépôts médiatisés reflète l'évolution de HAL vers un dispositif de suivi et d'évaluation de la recherche ;
  3. les laboratoires soutiennent majoritairement HAL, et ce soutien se concrétise – outre par la communication et l'incitation – par la création de collections et par l'utilisation de HAL pour les besoins des unités ;
  4. les études révèlent une diversité d'approches et de pratiques, et on peut identifier plusieurs types de laboratoires (*clusters*), par rapport à des variables quantitatives (nombre de dépôts, part d'auto-archivage avec texte intégral) et qualitatives (pratiques communautaires, actions d'engagement et de valorisation) ;
  5. une partie de la variation des pratiques individuelles et collectives est corrélée avec les disciplines et domaines scientifiques ;
  6. d'après les études, plusieurs variables déterminent les dépôts sur HAL (le nombre de chercheurs, les choix des laboratoires, les pratiques disciplinaires, la politique d'établissement et des pratiques individuelles) ; ces variables sont interdépendantes et font partie du même environnement.
- 29 Comme indiqué plus haut, nous allons présenter ici quelques résultats empiriques pour permettre une description plus détaillée des pratiques sur HAL de la part des laboratoires de recherche, en complément d'autres études et en particulier, des approches qualitatives.

## Méthodologie

- 30 L'échantillon du projet HAL/LO<sup>7</sup> est composé des laboratoires des dix grandes universités de recherche et membres de l'association Udice<sup>8</sup>, qui représentent 45 % des publications françaises (Observatoire des sciences et des techniques [OST], 2021)<sup>9</sup>.
- 31 À partir des sites des universités, nous avons identifié 1 246 laboratoires de recherche affiliés à une ou plusieurs de ces universités. Pour chaque unité, nous avons déterminé le nom, l'acronyme et/ou le numéro et le statut, et nous avons ajouté d'autres informations, notamment le domaine de recherche et la discipline, mais aussi les codes structures et le lien internet vers la collection sur HAL, à partir des

sites web de l'université et des unités, du répertoire national des structures de recherche (RNSR), du référentiel de la plateforme HAL (AuréHAL) et du moteur de recherche scanR.

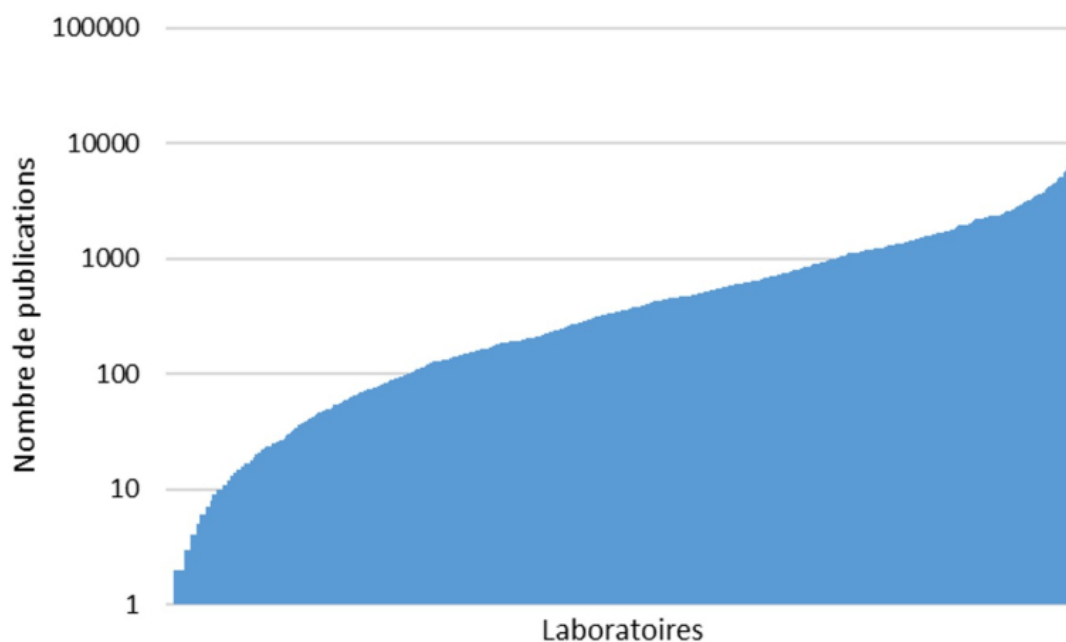
- 32 La composition de l'échantillon – le statut des laboratoires, leurs affiliations universitaires, leurs domaines et leurs disciplines – est décrite dans l'annexe.
- 33 L'extraction des données HAL a été effectuée via l'API de HAL (codes structures), à plusieurs reprises (mars-avril 2021). Les résultats ont été vérifiés, contrôlés et nettoyés par trois membres de l'équipe projet. L'analyse et la visualisation des données ont été réalisées avec Excel.

## Résultats

### L'utilisation de HAL

- 34 Est-ce que tous les laboratoires utilisent HAL ? Pour pratiquement tous les laboratoires, on trouve des publications dans HAL. Nous avons pu établir le nombre de publications présentes dans HAL pour 1 228 laboratoires (99 % de notre échantillon). Mais cela ne veut pas dire que les laboratoires utilisent HAL au sens propre du terme. Si pour certains laboratoires, le nombre de dépôts se chiffre par milliers, d'autres n'ont que très peu de publications.
- 35 En effet, le nombre de publications varie de 1 à 10 645, avec une moyenne de 843 et une médiane de 359 publications. La figure 1 montre le nombre des publications pour chaque laboratoire (échelle logarithmique).

Figure 1. Nombre de publications par laboratoire (N = 1 228 laboratoires)



- 36 Pour un quart des laboratoires, le nombre des publications dans HAL est inférieur ou égal à 100, pour un autre quart, ce nombre est égal ou supérieur à 1 000. Le rapport avec la taille des laboratoires est significatif mais, comme indiqué plus haut, n'explique pas tout.
- 37 Il ne s'agit pas d'une distribution de type Pareto (80/20) mais plutôt d'un début de longue traîne où 20 % des publications correspondent au dépôt de 70 % des laboratoires ; et où 20 % des laboratoires représentent ensemble seulement 67 % (et non pas 80 %) des dépôts.

## Typologie des dépôts

- 38 Combien de laboratoires déposent ? Quels types de documents ? Tous les types de documents et de données de HAL sont représentés dans les dépôts de ces laboratoires. Plus de la moitié des dépôts sont des articles de revue (52 %), suivis par des communications (23 %), des chapitres d'ouvrages (8 %) et des thèses (4 %). D'autres types de documents sont présents en moindre importance, comme des livres, des rapports, des *preprints* ou des posters (1-3 %) ou de façon marginale, comme des mémoires, des Habilitations à diriger des recherches (HDR), des présentations ou des brevets (<1%). Parmi les dé-

pôts des laboratoires se trouvent aussi quelques données de recherche, notamment des photos (1,1 %) et des vidéos (0,2 %).

- 39 Quand on regroupe les dépôts en fonction de la nature des documents et données, on obtient le tableau suivant : plus de trois dépôts sur cinq correspondent à des publications (articles, chapitres, livres écrits ou coordonnés), plus d'un tiers relève au moins partiellement de la littérature grise (communications, posters, présentations, thèses, mémoires, rapports, preprints, etc.), et 1 % correspond à différents types de données (images, vidéos, cartes, logiciels, etc.), (tableau 1).

**Tableau 1. Répartition des types de contenus (N = 1 035 612 dépôts)**

|  |      |
|--|------|
| Publications (articles, livres, etc.)                              | 63 % |
| Autres documents (communications, rapports, thèses, posters, etc.) | 36 % |
| Données (photos, vidéos, logiciels, etc.)                          | 1 %  |

- 40 L'article de revue est de loin le type de dépôt le plus important et universel au sens où pratiquement tous les laboratoires (99 %) ont déposé des articles dans HAL. Un peu moins de laboratoires – mais toujours un nombre élevé, huit sur dix voire davantage – ont déposé des communications, des thèses et des chapitres de livres. Un troisième groupe de documents a été déposé par 50-70 % de laboratoires : les *preprints*, les posters, les ouvrages écrits ou coordonnés et les rapports. Les autres types de documents ont été déposés par relativement peu de laboratoires. Le tableau 2 contient les valeurs pour ces deux variables.

**Tableau 2. L'importance des différents types de documents (N = 1 035 612 dépôts)**

|                  | Laboratoires | Dépôts |
|------------------|--------------|--------|
| Articles         | 99 %         | 52 %   |
| Communications   | 87 %         | 23 %   |
| Thèses           | 86 %         | 4 %    |
| Chapitres        | 80 %         | 8 %    |
| <i>Preprints</i> | 68 %         | 3 %    |

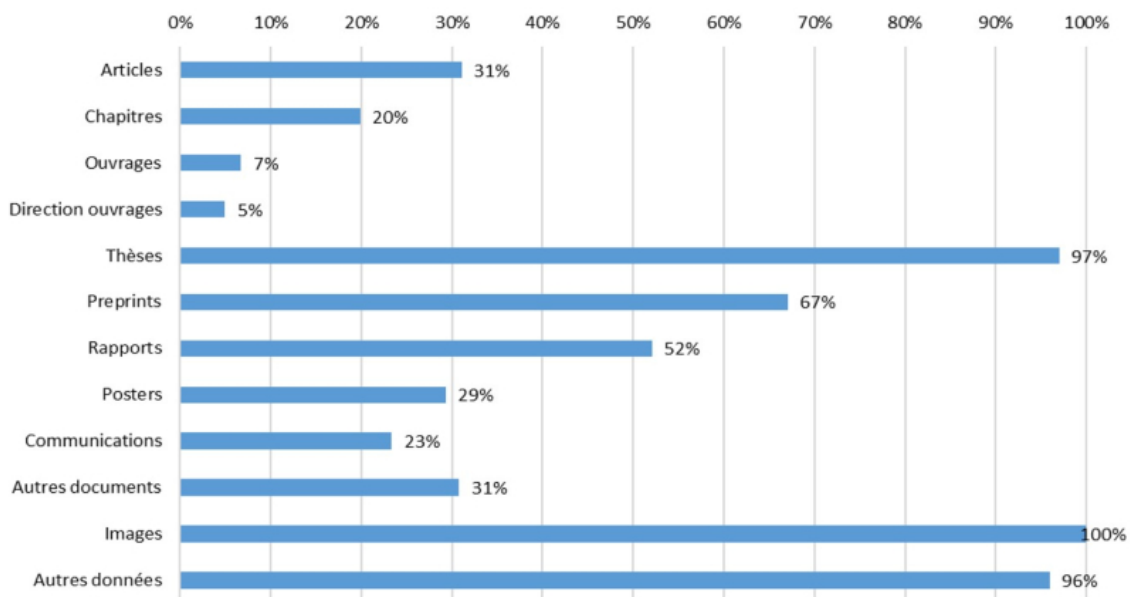
|                    |      |     |
|--------------------|------|-----|
| Posters            | 62 % | 1 % |
| Ouvrages           | 58 % | 2 % |
| Direction ouvrages | 55 % | 2 % |
| Rapports           | 51 % | 2 % |

- 41 En revanche, d'autres ressources ont été déposées par beaucoup moins de laboratoires, comme les brevets (18 %), des vidéos (16 %), des mémoires (12 %) ou encore des logiciels (5 %). Les laboratoires ne produisent pas les mêmes types de ressources et de résultats ; mais en même temps, leurs dépôts ne reflètent pas toujours toute la gamme de leur production d'une manière exhaustive.

## Documents en texte intégral

- 42 Combien de documents sont déposés en texte intégral ? Une troisième analyse porte sur le degré d'ouverture des différents types de dépôts, autrement dit, sur la part des dépôts avec fichiers (documents en texte intégral, données avec fichiers), laquelle correspond essentiellement à l'auto-archivage par les chercheurs.
- 43 Les publications – articles et livres – ont un degré d'ouverture relativement bas, en dessous de 31 % (la moyenne pour l'ensemble des dépôts est 32 %, voir plus loin). Ce constat est identique pour les communications et les posters, ainsi pour l'ensemble des autres documents non publiés (HDR, mémoires, etc.). *A contrario*, 52 % des rapports et 67 % des *preprints* sont librement accessibles sur HAL, ainsi que pratiquement toutes les thèses (figure 2).

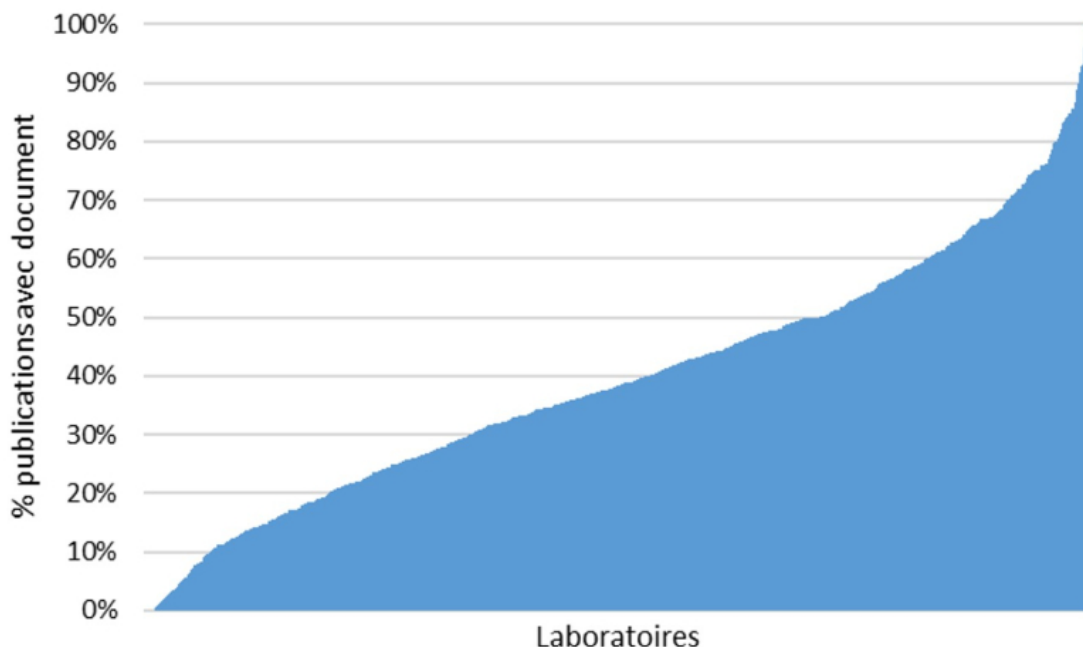
**Figure 2. La part des dépôts avec fichiers (% *open access*) (N = 1 035 612 dépôts)**



- 44 La figure 2 contient à titre comparatif également le pourcentage des données déposées sur HAL, en particulier les images (photos). Comme pour les thèses, pratiquement toutes les données sont librement accessibles. L'explication est identique : pour les thèses, ainsi que pour les données, le dépôt des fichiers est obligatoire ; il est quasiment impossible de signaler une thèse, une vidéo ou une photo sur HAL uniquement avec une notice (métadonnées), sans déposer le (ou les) fichier(s).
- 45 Seulement une partie des publications dans HAL est déposée avec le document, tandis que les autres publications correspondent à des notices bibliographiques (métadonnées) sans document. Au moment de la rédaction de l'article (avril 2023), le taux de publications avec document est de 26 % pour l'ensemble des dépôts HAL. Pour les publications de l'échantillon des 1 228 laboratoires, ce taux est de 32 %, donc supérieur à la moyenne de HAL. La médiane se situe à 38 % – autrement dit, la moitié des laboratoires analysés a un « taux d'ouverture » (= des publications librement accessibles sur HAL) d'au moins 38 %. Ce taux varie de 1 % à 100 % ; les extrêmes correspondent surtout à des laboratoires ayant très peu de publications sur HAL (figure 3).



Figure 3. Le taux de publications avec document par laboratoire (N = 1 228)



- 46 Nous trouvons aussi plusieurs laboratoires avec des centaines, voire des milliers de publications dans HAL mais seulement 1 % ou 2 % de documents (plein texte), tout comme nous identifions quelques laboratoires avec beaucoup de publications, et en même temps, beaucoup de documents (égal ou supérieur à 80 %).
- 47 Ajoutons ici déjà que ce « taux d'ouverture » ne correspond pas au pourcentage de toutes les publications librement accessibles sur le web – ce pourcentage est nécessairement supérieur au taux d'ouverture sur HAL, car pour un certain nombre de publications, HAL contient uniquement la notice de signalement, tandis que le document (en texte intégral) est librement accessible sur une autre plateforme de revues ou de livres, comme OpenEdition, ScienceDirect, Springer Link ou MDPI<sup>10</sup>.

## Collections

- 48 Combien de laboratoires ont créé une collection sur HAL ? HAL contient un grand nombre de portails institutionnels et de collections, et ce nombre augmente continuellement. Ainsi, le bilan de 2018 indique que HAL rassemble « 6 547 collections construites à partir de

critères variés : collections de laboratoires, projets de l'Agence nationale de la recherche (ANR) ou européens, congrès », dont 842 nouvelles créations<sup>11</sup>. Le rapport d'activité 2019 du CCSD évoque la création de 716 nouvelles collections, dont 341 collections de laboratoires<sup>12</sup>.

- 49 En 2021 et pour un échantillon de 1 246 laboratoires, nous avons identifié une collection HAL pour 626 unités (50 %). Autrement dit, la moitié des laboratoires a pris la décision de créer une collection. Par rapport aux domaines scientifiques et au statut des unités, nous avons pu constater deux particularités significatives : les laboratoires en sciences de la vie et santé ont créé moins de collections (seulement 29 %) qu'en sciences et technologies (64 %) et en SHS (60 %) ; et la part des laboratoires mixtes (unité mixte de recherche, UMR) avec une collection (61 %) est supérieure aux autres (37 %).

## Discussion

- 50 Pour au moins trois raisons, les résultats de notre étude paraissent représentatifs pour le contenu de HAL : les dépôts analysés (1 035 612) correspondent à environ un tiers du contenu de HAL au moment de l'extraction en 2021 ; leur répartition (typologie, voir tableau 2) est assez similaire aux dépôts de HAL par les chercheurs du CNRS (avril 2022)<sup>13</sup> ; et d'une manière plus générale, comme indiqué plus haut, la production de ces laboratoires représente une partie significative de la recherche française (voir Observatoire des sciences et des techniques [OST], 2021).
- 51 Cependant, plusieurs raisons limitent la portée des résultats. En l'absence d'un identifiant fiable pour les structures et organisations scientifiques, il est difficile d'identifier et de décrire avec précision et d'une manière exhaustive les unités de recherche d'une université<sup>14</sup>. Malgré plusieurs contrôles, l'échantillon peut contenir quelques doublons ou des unités fermées, avec des descriptions imprécises ou erronées. Par ailleurs, nous avons écarté les structures fédératives. Du fait de la méthode d'extraction basée sur les identifiants de structure, le corpus des dépôts (publications) contient nécessairement des doublons que nous n'avons pas éliminés, car notre intérêt portait avant tout sur l'activité de chaque laboratoire.

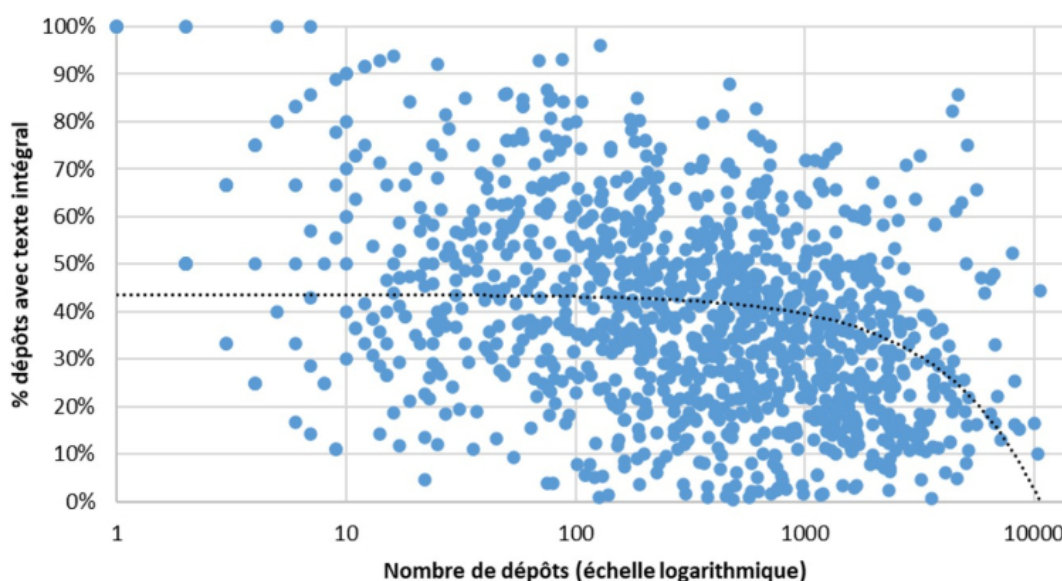
## Le rôle des laboratoires

- 52 Quel est le rôle des laboratoires de recherche dans l'environnement de la science ouverte ? Les résultats de notre analyse quantitative confirment que HAL contient des publications de tous les laboratoires. Tous les laboratoires sont présents sur HAL, d'une manière ou d'une autre, avec beaucoup ou peu de publications. Cela va dans le sens des résultats de l'enquête qualitative : « (presque) tous connaissent et utilisent HAL, plus ou moins [...] HAL est sorti de sa phase "de jeunesse" pour devenir un partenaire obligé et essentiel des chercheurs et de leurs unités de recherche » (Schöpfel *et al.*, 2021b).
- 53 En parallèle, nos résultats confirment le constat de diversité décrit par des études et des enquêtes antérieures. Au lieu de parler d'un rôle spécifique et homogène de tous les laboratoires, il convient de décrire un large éventail de situations et de pratiques hétérogènes, et ceci à plusieurs niveaux, comme par rapport au nombre de dépôts, au pourcentage des documents en plein texte, à la typologie des documents déposés, à la création d'une collection ou encore aux modalités d'alimentation. Nous avons déjà évoqué plusieurs facteurs potentiels pour expliquer cette variation, comme la discipline et l'existence d'une culture communautaire, le statut de l'unité et sa ou ses tutelles, ou encore les ressources humaines et matérielles, y compris la présence d'un poste de documentaliste ou d'une base de données interne.
- 54 L'indicateur le plus visible d'une pratique institutionnelle assumée est la présence d'une collection sur HAL. En 2021, la moitié des laboratoires ont pris l'initiative d'une telle collection. À l'instar de Lille (Kergosien & Schöpfel, 2021), leur nombre augmente. Il est plus que probable qu'au moment de la rédaction de cet article (2023), ce pourcentage soit significativement plus élevé. Mais que dire des autres laboratoires, ceux qui n'utilisent pas HAL pour signaler et suivre leur production scientifique ? S'agit-il d'unités de recherche universitaires, sans personnel CNRS ? Ou de laboratoires dans certains domaines particuliers qui ont leurs propres outils reconnus et indispensables, comme PubMed Central en médecine ou RePEc en sciences économiques ? En tout cas, les différences persistent.

55 Deux autres indicateurs pour caractériser les pratiques et l'utilisation de HAL par les laboratoires sont le nombre total de dépôts et la part des dépôts avec texte intégral (figure 4). En comparaison avec l'analyse des laboratoires de Lille (Schöpfel, 2020 ; Kergosien & Schöpfel, 2021), on retrouve les mêmes groupes de laboratoires :

1. utilisation de HAL comme une vitrine pour la production scientifique (une collection avec beaucoup de publications ; un taux élevé de documents en texte intégral) ;
2. utilisation de HAL comme un gestionnaire de références (beaucoup de dépôts ; peu de documents en texte intégral) ;
3. pas de stratégie (pas de collection ; très peu de dépôts).

**Figure 4. Nombre de dépôts et pourcentage des dépôts avec texte intégral (N = 1 228)**



56 En appliquant les mêmes critères que pour l'étude à Lille<sup>15</sup>, on obtient 19 % pour le *cluster* 1 (vitrine), 10 % pour le *cluster* 2 (gestionnaire) et 26 % pour le *cluster* 3 (absence de stratégie).

57 Deux observations paraissent intéressantes. D'une part, en comparaison aux résultats de Lille, l'importance de ces *clusters* diminue ; ici, ils ne représentent plus que 55 % des laboratoires, bien moins que les 96 % à Lille en 2019 (Schöpfel, 2020). Au lieu de décrire quelques cas types (*clusters*), on devrait plutôt parler d'une longue traîne qu'on

pourrait caractériser à partir des données quantitatives de deux manières :

- les « meilleurs » laboratoires (20 % des laboratoires avec au total 67 % des publications) : la médiane de la part des dépôts avec texte intégral est seulement de 26 %. Dans ce groupe, nous trouvons les laboratoires avec beaucoup de publications sur HAL, mais moins de dépôts en texte intégral ;
- la « longue traîne » au sens propre (70 % des laboratoires avec au total 20 % des publications) : la médiane de la part des dépôts en texte intégral est 43 %. On trouve ici tous les laboratoires avec moins de dépôts de publications, mais une part plus élevée de documents librement accessibles.

58 La courbe de tendance (figure 4) illustre cette observation : les laboratoires avec peu de publications sur HAL ont en moyenne davantage de dépôts en texte intégral que ceux avec beaucoup de publications. C'est un constat qui rejoint une observation concernant les archives institutionnelles : à partir du moment où les institutions ont saisi l'opportunité offerte par les archives institutionnelles pour prendre le contrôle de leur propre production scientifique, l'exhaustivité du contenu devient plus importante que l'accès au texte intégral. « Les grandes archives exhaustives permettent une évaluation scientométrique des résultats et de la productivité de la recherche ; dans ce cas, les métadonnées sont importantes, tandis que l'accès au texte intégral est secondaire, marginal » (Prost & Schöpfel, 2014). Nous y reviendrons plus loin.

59 La deuxième observation concerne les laboratoires avec un nombre élevé de publications dans HAL (>1 000) qui ont tendanciellement moins de documents en texte intégral et davantage de notices (figure 4). Néanmoins, pour ces laboratoires, nous pouvons différencier deux groupes :

- 26 laboratoires avec beaucoup de dépôts en texte intégral (>60%). Il s'agit de laboratoires notamment dans les disciplines de l'informatique, des mathématiques et de la physique, avec des communautés acquises à la science ouverte et une tradition forte de dépôt sur HAL (voie verte) (voir Baruch, 2007 ; Mahé, 2017) ;
- 105 laboratoires avec peu de dépôts en texte intégral (<20%). On trouve ici davantage de laboratoires dans les SHS mais aussi en biologie, en chimie ou en génie civil, pour lesquels le signalement de leur production semble plus important que la communication directe des publications.

60 Nous observons donc de nouveau l'effet de deux variables, celle des disciplines avec leurs propres outils, traditions et pratiques, et celle du choix institutionnel concernant l'utilisation de HAL comme un outil de suivi et d'évaluation.

## L'impact des disciplines scientifiques

61 Nous avons évoqué à plusieurs reprises le lien entre l'appartenance à un domaine ou une discipline ou une communauté particulière, et la diversité des pratiques en matière d'information scientifique et de l'utilisation de la plateforme HAL, que ce soit la création de collections, l'auto-archivage ou l'alimentation par d'autres moyens, ou d'autres actions. Ces différences s'ajoutent à d'autres variables dont elles modèrent l'impact, comme notamment le statut des unités et des personnels (impact de la politique du CNRS) et la taille des unités (corrélation avec le nombre de dépôts).

62 Toutes les variables de notre analyse quantitative sont impactées d'une manière significative<sup>16</sup> ; voici une synthèse des résultats :

63 **Utilisation de HAL** : les laboratoires en biologie, médecine et santé déposent en moyenne moins que les autres ; quatre à cinq fois moins par exemple qu'en informatique, mathématique ou physique (tableau 3).

**Tableau 3. Nombre moyen de dépôts par laboratoire et discipline**<sup>17</sup>

| Discipline   | Moyenne |
|--|---------|
| Sciences et technologies de l'information et de la communication | 1 521   |
| Mathématiques et leurs interactions                              | 1 496   |
| Physique   | 1 313   |
| Sciences pour l'ingénieur  | 1 177   |
| Sciences de la terre, de l'univers et de l'espace                | 1 086   |
| Sciences de la société   | 980     |
| Chimie   | 972     |
| Sciences agronomiques et écologiques                             | 915     |
| Sciences humaines et humanités                                   | 787     |
| Biologie, médecine, santé  | 306     |

- 64 Le tableau montre également que les dépôts en SHS sont en moyenne inférieurs aux dépôts en sciences et technologies, à l'exception de la chimie et de l'agronomie.
- 65 **Typologie des dépôts** : pour presque tous les laboratoires on trouve des articles de revues dans HAL (voir tableau 2). Aussi, on ne constate pas de différence par rapport au dépôt des thèses (80-90 % de laboratoires dans tous les domaines). L'impact des domaines est par contre visible et significatif pour les autres types de documents : autour de 90 % des laboratoires en SHS et en droit, économie et gestion déposent des livres, contre environ 60 % en sciences et technologies et seulement environ 20 % en sciences de la vie et santé. Dans ce dernier domaine, significativement moins de laboratoires déposent des chapitres, des rapports, des *preprints* et même des communications (tableau 4).

**Tableau 4. Dépôts par type de document et domaine (en % des 1 246 laboratoires) <sup>18</sup>**

|           | Nb laboratoires | Communications | Thèses | Chapitres | Pre-prints | Ouvrages | Direction ouvrages | Rapports |
|-----------|-----------------|----------------|--------|-----------|------------|----------|--------------------|----------|
| ALLSHS    | 301             | 94 %           | 84 %   | 93 %      | 75 %       | 89 %     | 86 %               | 61 %     |
| DEG       | 85              | 89 %           | 80 %   | 95 %      | 78 %       | 88 %     | 87 %               | 72 %     |
| SDV santé | 415             | 69 %           | 79 %   | 58 %      | 46 %       | 24 %     | 20 %               | 32 %     |
| ST        | 445             | 94 %           | 91 %   | 86 %      | 80 %       | 62 %     | 58 %               | 56 %     |
| Total     | 1 246           | 87 %           | 86 %   | 80 %      | 68 %       | 58 %     | 55 %               | 51 %     |

- 66 **Documents en texte intégral** : les différences par domaine sont également significatives au niveau des dépôts en texte intégral (tableau 5).

**Tableau 5. La part des dépôts en texte intégral, par domaine (1 031 155 dépôts, sans les annexes)**

|           |      |
|-----------|------|
| ALLSHS    | 28 % |
| DEG       | 19 % |
| SDV santé | 37 % |

|       |      |
|-------|------|
| ST    | 35 % |
| Total | 32 % |

- 67 La part des dépôts en texte intégral en droit, économie et gestion (19 %) est inférieure à la moyenne de tous les dépôts (32 %), tout comme dans une moindre mesure en SHS (28 %).
- 68 **Collections** : même constat pour la création des collections – il y a des différences significatives au niveau des domaines scientifiques, avec davantage de collections pour les laboratoires en SHS (60 %) et en sciences et technologies (64 %) qu'en droit, économie et gestion (47 %) et surtout en sciences de la vie et santé (29 %) (tableau 6).

**Tableau 6. La part des collections de laboratoire, par domaine (1 246 laboratoires)**

|           |      |
|-----------|------|
| ALLSHS    | 60 % |
| DEG       | 47 % |
| SDV santé | 29 % |
| ST        | 64 % |
| Total     | 50 % |

- 69 Ces quelques chiffres reflètent une réalité diversifiée et confirment les constats antérieurs : les laboratoires et leurs chercheurs n'utilisent pas HAL de la même manière, que ce soit par rapport à la création d'une collection, au dépôt de certains types de documents ou à l'auto-archivage des documents en texte intégral, ou tout simplement par rapport au nombre des dépôts. Les grands domaines et les disciplines jouent un rôle significatif ; mais ils sont une sorte d'abstraction pour une grande variété de comportements de publications, de tutelles, d'équipements et de ressources mais aussi d'outils et stratégies dans l'environnement de la science ouverte.
- 70 Ajoutons aussi qu'une analyse plus fine s'impose dans la mesure où une partie importante des laboratoires (29 %) couvrent plusieurs disciplines, avec des équipes multidisciplinaires et/ou des équipes appartenant à des disciplines différentes. Autrement dit, il y a très certainement des laboratoires dont les équipes fonctionnent différemment par rapport à HAL ; une diversité qui se superpose avec les inci-



tations et obligations différentes selon l'affiliation et le statut des personnels en question.

## L'évolution du dispositif HAL

- 71 HAL a été créé sur le modèle d'arXiv, comme « une archive ouverte de texte intégral (afin de) réaliser un arXiv multidisciplinaire » (Charnay, 2019), avec une « culture de dépôt » basée sur l'auto-archivage de documents (*preprints*, articles, etc.) en texte intégral par les auteurs. Quelques années après le lancement, plusieurs études montrent que l'objectif initial n'a pas été atteint ; certaines disciplines sont surreprésentées (physique, informatiques, SHS) par rapport à d'autres (chimie, biologie, médecine) ; une part importante des dépôts ne contient pas de texte intégral et/ou n'est pas réalisée par les auteurs ; les dépôts ne représentent qu'une partie de la production scientifique de la France ; et tous les dépôts ne sont pas accompagnés par des métadonnées suffisamment renseignées.
- 72 Comme Daniel Charnay, on peut regretter cette « dérive bibliométrique des archives ouvertes » et déplorer une situation « triste », « catastrophique » ou « ubuesque » qui s'éloigne de « l'esprit original » de HAL (Charnay, 2019). Force est de constater, cependant, qu'il ne s'agit pas d'un phénomène isolé, mais d'une évolution générale qui a été décrite comme un rapprochement des archives institutionnelles et des systèmes d'information recherche (Schöpfel & Azeroual, 2021 ; Schöpfel *et al.*, 2023) qui eux permettent « l'agrégation, la conservation et l'utilisation de métadonnées sur les activités de recherche » (Bryant *et al.*, 2017), afin de produire des connaissances fiables sur la recherche et d'aider les institutions scientifiques à assurer le suivi des projets scientifiques et à fournir des informations nécessaires aux bilans, aux rapports d'activité et aux campagnes d'évaluation (De Castro, 2018).
- 73 Depuis plusieurs années, on peut observer une convergence progressive de ces deux types de systèmes (= systèmes de diffusion et systèmes d'évaluation), par le biais d'un transfert systématique de métadonnées, par des zones de chevauchement dans les tâches que les deux types de systèmes accomplissent, ou en permettant à l'un d'entre eux de reprendre les caractéristiques de l'autre, offrant ainsi une fonctionnalité intégrée à un seul système (De Castro *et al.*, 2014).

- 74 Concrètement, cela signifie que les archives ouvertes développent de plus en plus leurs fonctionnalités d'évaluation et de suivi, tandis que les systèmes d'information recherchent et intègrent progressivement des fonctionnalités d'archives ouvertes. L'évolution de HAL n'est donc pas une exception mais plutôt une particularité, dans la mesure où, pour une archive ouverte, ce développement se situe généralement au niveau d'une institution (université, institut de recherche, etc.) et pas, comme pour HAL, au niveau d'un État. Le plus souvent, ce développement répond à la demande et aux besoins d'une institution ; quant à HAL, cette demande émane aussi bien des établissements, organismes et laboratoires de recherche que de l'administration centrale de l'État qui souhaite disposer d'un « outil pour recenser l'exhaustivité de la production scientifique française » (Charnay, 2019). Il s'agit là d'une « rupture conceptuelle importante dans l'histoire de l'archive » qui se traduit, par exemple (mais pas seulement), par le déploiement de « mécanismes automatisés de collecte ciblée des publications des auteurs-chercheurs utilisant HAL (pour) automatiser la récupération de publications à la demande des chercheurs (et pour) leur proposer des publications dont nous pensons qu'ils sont les auteurs et qui sont présentes sur des plateformes extérieures (celles d'éditeurs ou serveurs de preprint par exemple) » (Berthaud *et al.*, 2021).
- 75 Cette nouvelle stratégie de HAL, la création de références sans document par les laboratoires, mais aussi l'incitation forte par le CNRS au dépôt de la production scientifique par les chercheurs, tout cela reflète cette recherche d'exhaustivité et accélère l'évolution de HAL d'une archive ouverte de publications, dans le cadre d'une communication scientifique directe, vers un outil bibliométrique d'évaluation et de suivi de la production scientifique des chercheurs et des institutions.
- 76 En l'absence d'autres instruments d'évaluation, tout un écosystème d'outils a été créé, des outils qui exploitent les données de HAL pour des besoins des laboratoires, des universités et de l'État. Mais en conséquence, d'autres questions se posent pour HAL, comme fournisseur de données et dispositif central de l'évaluation de la recherche : des questions concernant soit la provenance et la qualité des données, soit l'interopérabilité du modèle de données avec d'autres systèmes d'information recherche, soit les fonctionnalités,

sans oublier les questions relatives à la professionnalisation documentaire des déposants (contributeurs) et aux aspects éthiques de l'utilisation de HAL.

## Conclusion

- 77 Notre analyse quantitative des dépôts effectués par les laboratoires de recherche des universités Udice ajoute de l'évidence empirique aux analyses antérieures et notamment aux enquêtes qualitatives, en confirmant la généralisation (ou banalisation) de l'utilisation de HAL par les laboratoires, un paysage contrasté avec certaines approches caractéristiques et le développement d'une longue traîne, des différences disciplinaires et l'évolution de HAL d'un dispositif à usage individuel (auto-archivage) vers un dispositif à destination des institutions (suivi, évaluation), dans lequel les deux principes – la communication directe par auto-archivage (voie verte du libre accès) et le suivi de la production scientifique – coexistent.
- 78 Pour approfondir la compréhension de la réalité sur le terrain, nous sommes en train d'analyser les résultats d'interviews de 50 laboratoires. Il s'agit en particulier de mieux contextualiser et d'illustrer certaines variables à l'origine de la diversité des situations et pratiques – les ressources humaines, les outils, le soutien institutionnel, le statut des personnels, les communautés disciplinaires, les équipes, etc., mais aussi le rôle des documentalistes et l'impact d'un engagement individuel.
- 79 Pour la suite, il serait intéressant de mener une analyse longitudinale sur deux ou trois ans, afin d'étudier l'évolution des pratiques des laboratoires dans la durée, notamment par rapport à l'impact des politiques publiques en faveur de la science ouverte, sous l'aspect des différences disciplinaires et communautaires. On pourrait également observer les répercussions de la mise en œuvre de la refonte de l'ergonomie et des services cœur de HAL (projet EquipeX HALiance)<sup>19</sup>.

## Annexes

### Le statut des 1 246 laboratoires

|        |  |
|--------|--|
| UMR    | 689  |
| UR/EA  | 336  |
| Autres | 221 (services, départements, instituts, USR, etc.) |

## L'affiliation universitaire des laboratoires

|                        |     |
|------------------------|-----|
| Aix-Marseille          | 105 |
| Bordeaux               | 75  |
| Côte d'Azur            | 42  |
| Grenoble Alpes         | 99  |
| Lyon 1                 | 143 |
| Strasbourg             | 76  |
| Paris Cité             | 220 |
| Paris Saclay           | 228 |
| Paris Sciences Lettres | 123 |
| Sorbonne Université    | 135 |

## Les domaines des laboratoires

|                     |     |
|---------------------|-----|
| ALLSHS              | 301 |
| DEG                 | 85  |
| Science technologie | 445 |
| SDV santé           | 415 |

## Les disciplines des laboratoires (RNSR)

|                                      |     |
|--------------------------------------|-----|
| Biologie, médecine et santé          | 482 |
| Chimie                               | 139 |
| Mathématiques et leurs interactions  | 58  |
| Physique                             | 150 |
| Sciences agronomiques et écologiques | 80  |
| Sciences de la société               | 203 |

|  |     |
|--|-----|
| Sciences de la terre, de l'univers et de l'espace                | 79  |
| Sciences et technologies de l'information et de la communication | 121 |
| Sciences humaines et humanités                                   | 282 |
| Sciences pour l'ingénieur  | 165 |

## BIBLIOGRAPHIE

---

Baruch, P. (2007). Open access developments in France: the HAL Open Archives System. *Learned Publishing*, 20 (4), 267-282. DOI : <https://doi.org/doi:10.1087/095315107x239636>.

Berthaud, C., Charnay, D., Fargier, N. (2021). Diffuser et pérenniser le savoir scientifique : 20 ans d'histoire de HAL. *Histoire de la recherche contemporaine*, 10 (2). DOI : <https://doi.org/10.4000/hrc.6330>.

Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., Wright, J. (2017). *Research Information Management: Defining RIM and the Library's Role*. OCLC, Dublin, OH. Repéré à : <https://www.oclc.org/research/publications/2017/oclcresearch-defining-rim.html>.

Charnay, D. (2019). Avec HAL, nous voulions créer un arXiv multidisciplinaire. *Hermès*, 85 (3), 94. DOI : <https://doi.org/10.3917/herm.085.0094>

Couperin (2020). *Enquête sur les archives ouvertes françaises menées par Couperin en 2019. Résultats et analyses*. Rapports et études Couperin n° 2. Paris : Consortium Couperin. Repéré à : <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>.

Deboin, M.-C. (2015). *Déposer ses publications dans une archive ouverte, en 8 points*. Montpellier : CIRAD. DOI: <https://doi.org/10.18167/coopist/0037>.

De Castro, P. (2018). *Mapping the European CRIS infrastructure and its potential applications*. Communication présentée à Antwerp ECOOM Workshop "Working with National Bibliographic Databases for Research Output", 10 au 11 septembre 2018, Anvers. Repéré à : <https://dSPACECRIS.eurocris.org/handle/11366/705>.

De Castro, P., Shearer, K., Summann, F. (2014). *The Gradual Merging of Repository and CRIS Solutions to Meet Institutional Research Information Management Requirements*. Communication présentée à CRIS2014, 12<sup>th</sup> International Conference on Current Research Information Systems, 13 au 15 mai 2014, Rome. DOI : <https://doi.org/doi:10.1016/j.procs.2014.06.007>.

Gayoso, E. (2020). *La diffusion sur Hal, Academia et ResearchGate des articles de recherche des revues françaises de Sciences Humaines et Sociales*. Paris : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. Repéré à : <https://www.enseignementsup-recherche.gouv.fr/cid136723/le-soutien-a-l-edition-scientifique.html>.

- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Hilf, E. R. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review*, 30 (4), 310-314. DOI : <https://doi.org/10.1016/j.serrev.2004.09.013>.
- Kergosien, E., Schöpfel, J. (2021). *L'usage de la plateforme HAL par les unités de recherche de l'Université de Lille. La situation en 2021*. Preprint. Repéré à : <https://hal.science/hal-03372596v1>.
- Lafon, S. (2013). *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*. Toulouse : université de Toulouse 3. Repéré à : [https://hal.archives-ouvertes.fr/sic\\_00977548](https://hal.archives-ouvertes.fr/sic_00977548).
- Larrieu, M., Schöpfel, J. (2022). *Éclairer les différences disciplinaires dans le contexte science ouverte. Analyse avec les publications de l'archive ouverte HAL*. Communication présentée à DocSoc 2022, 8<sup>e</sup> Conférence Document numérique & Société, du 23 au 24 juin 2022, Liège. Repéré à : <https://hal.univ-lille.fr/hal-03760316v1>.
- Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Portal: Libraries and the Academy*, 3 (2), 327-336. DOI : <https://doi.org/10.1353/pla.2003.0039>.
- Magron A. (2017). Utiliser les archives ouvertes pour valoriser ses travaux l'exemple de HAL-SHS. Dans Schnedeker, C., Aleksandrova, A. (dir.), *Le document en France : mode(s) d'emploi (227-239)*. Berlin: Peter Lang. Repéré à : [https://hal.archives-ouvertes.fr/sic\\_01697164](https://hal.archives-ouvertes.fr/sic_01697164).
- Mahé, A. (2017). Les pratiques informationnelles des chercheurs dans l'enseignement supérieur et la recherche : regards sur la décennie 2000-2010. Dans Chartron, G., Epron, B., Mahé, A. (dir.), *Pratiques documentaires numériques à l'université (12-41)*. Villeurbanne : Presses de l'Enssib. DOI : <https://doi.org/10.4000/books.pressesensib.1171>.
- Mahé, A., Prime-Claverie, C. (2017). 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. DOI : <https://doi.org/10.4000/rfsic.3315>.
- Morais, R., Saenen, B., Garbuglia, F., Gaillard, V. (2021). *From Principles to Practices: Open Science at Europe's Universities - 2020-2021 EUA Open Science Survey Results*. Bruxelles, European University Association. <https://www.eua.eu/resources/publications/976:from-principles-to-practices-open-science-at-europe-s-universities-2020-2021-eua-open-science-survey-results.html>.
- Neugebauer, T., Murray, A. (2013). The Critical Role of Institutional Services in Open Access Advocacy. *International Journal of Digital Curation*, 8 (1), 84-106. DOI : <https://doi.org/10.2218/ijdc.v8i1.238>.
- Observatoire des sciences et des techniques (OST). (2021). *La position scientifique de la France dans le monde et en Europe, 2005-2018*. Paris : Hcéres – Observatoire des sciences et des tech-

- Reperé à : <https://www.hceres.fr/fr/actualites/parmi-les-pays-de-recherche-intensive-la-france-conserve-un-profil-scientifique>.
- Prime-Claverie, C., Mahé, A. (2013). Sites de dépôt en libre accès et formes de médiations : quelles évolutions ? Dans Boustany, J., Broudoux, É., Chartron G. (dir.), *La médiation numérique : renouvellement et diversification des pratiques. Actes du colloque Document numérique et société, Zagreb 2013* (125-139). Bruxelles : De Boeck, ADBS. DOI : <https://doi.org/10.3917/dbu.chron.2013.01.0125>.
- Prost, H., Schöpfel, J. (2014). Degrees of Openness: Access Restrictions in Institutional Repositories. *D-Lib Magazine*, 20 (7/8). DOI : <https://doi.org/10.1045/july2014-prost>.
- Schöpfel, J. (2020). L'usage de la plateforme HAL par des unités de recherche. Le cas de l'université de Lille. *I2D - Information, Données & Documents*, 3 (3), 167-198. DOI : <https://doi.org/10.3917/i2d.203.0167>.
- Schöpfel, J., Prost, H. (2010). *Développement et usage des archives ouvertes en France. Rapport. 1<sup>re</sup> partie : développement*. Villeneuve d'Ascq : Université de Lille 3. [http://archive-sic.ccsd.cnrs.fr/sic\\_00497389/fr/](http://archive-sic.ccsd.cnrs.fr/sic_00497389/fr/).
- Schöpfel, J., Ferrant, C., André, F., Fabre, R. (2016). Ready for the Future? À Survey on Open Access with Scientists from the French National Research Center (CNRS). *Interlending & Document Supply*, 44 (4), 141-149. DOI : <https://doi.org/doi:10.1108/ILDS-06-2016-0023>.
- Schöpfel, J., Prost, H., Fraisse, A., Chaudiron, S. (2018a). *Valoriser les publications d'un laboratoire universitaire dans l'environnement de la science ouverte : retour d'expérience de la collection GERriCO sur HAL*. Communication présentée à ICOA 2018, 3<sup>e</sup> Colloque international sur le libre accès, 28 au 30 novembre 2018, Rabat. <https://hal.archives-ouvertes.fr/hal-01940352>.
- Schöpfel, J., Ferrant, C., André, F., Fabre, R. (2018b). Research Data Management in the French National Research Center (CNRS). *Data Technologies and Applications*, 52 (2), 248-265. DOI : <https://doi.org/doi:10.1108/DTA-01-2017-0005>.
- Schöpfel, J., Prost, H., Ndiaye, I. E. H. (2019). *Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL*. Communication présentée à GL 21, International Conference on Grey Literature, du 22 au 23 octobre 2019, Hanovre. Reperé à : <https://hal.archives-ouvertes.fr/hal-02300017>.
- Schöpfel, J., Azeroual, O. (2021). Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger. Dans Baker, D., Ellis, L. (dir.), *Future Directions in Digital Information. Predictions, Practice, Participation* (19-37). Oxford: Chandos Publishing. Reperé à : <https://hal.univ-lille.fr/hal-02994300v1>.
- Schöpfel, J., Kergosien, E., Prost, H., Thiault, F. (2020). *The Grey Side of the Green Road. Empirical Assessment of Academic Publishing in the HAL Open Repository*. Communication présentée à GL 2020. 22<sup>nd</sup> International Conference on Grey Literature "Applications of Grey Literature for Science and Socie-

ty", 19 novembre 2020, Rome. Repéré à : <http://greyguide.isti.cnr.it/attachments/article/51/GL2020%20Conference%20Proceedings.pdf#page=45>.

Schöpfel, J., Kergosien, E., Prost, H., Thiault, F. (2021). *Grey Literature in Open Repositories: New Insights and New Issues*. Communication présentée à GL 23, International Conference on Grey Literature, du 6 au 7 décembre 2021, Amsterdam. DOI : <https://hal.univ-lille.fr/hal-03604697>.

Schöpfel, J., Kergosien, E., Prost, H., Barrié, J. (2022). « Pas si simple que ça... » : une enquête sur l'usage de HAL par les unités de recherche des universités IdEx. *I2D - Information, Données & Documents*, 2 (2), 150-183. DOI : <https://doi.org/10.3917/i2d.222.0150>.

Schöpfel, J., Chaudiron, S., Jacquemin, B., Kergosien, É., Prost, H., Thiault, F. (2023). *The Transformation of the Green Road to Open Access*. *Preprints.Org*, 2023020268. DOI : <https://doi.org/10.20944/preprints202302.0268.v1>.

Tabariès, A. (2022). Vers une métrique pour évaluer les métadonnées de documents scientifiques. *Revue française des sciences de l'information et de la communication*, 24. DOI : <https://doi.org/10.4000/rfsic.12258>.

Westell, M. (2006). Institutional Repositories: Proposed Indicators of Success. *Library Hi Tech*, 24 (2), 211-226. DOI : <https://doi.org/10.1108/07378830610669583>.

Yeates, R. (2003). Institutional Repositories. *Vine*, 33 (2), 96-101. DOI : <https://doi.org/10.1108/03055720310509064>.

## NOTES

---

1 Voir Deuxième Plan national pour la science ouverte 2021-2024. En ligne : <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525>.

2 Projet HAL/LO Valorisation sur HAL de la production des laboratoires dans l'environnement de la science ouverte. En ligne : <http://gis-reseau-urfist.fr/hal-lo-valorisation-sur-hal-de-la-production-des-laboratoires-dans-lenvironnement-de-la-science-ouverte/>.

3 OpenDOAR : <https://v2.sherpa.ac.uk/opendoar/> (11 avril 2023).

4 HAL : <https://hal.science/>.

5 Voir OpenDOAR. En ligne : <https://v2.sherpa.ac.uk/id/repository/166>.

6 CCSD, Rapport d'activité, 2022. En ligne : <https://www.ccsd.cnrs.fr/2023/03/ccsd-publication-du-rapport-d-activite-2022/>.

7 Voir la note 1.



8 Udice : Aix-Marseille, Bordeaux, Côte d'Azur, Grenoble-Alpes, Lyon 1, Paris Cité, Paris-Saclay, Paris Sciences et Lettres, Sorbonne Université, Strasbourg. En ligne : <https://www.udice.org/a-propos/>.

9 D'après le site de l'Udice, les dix universités représentent 33 800 enseignants-chercheurs et chercheurs, 24 000 doctorants (soit près de 50 % des doctorants français), 602 millions d'euros de contrats de recherche et 2/3 des publications françaises les plus citées au monde. En ligne : <https://www.udice.org/a-propos/>.

10 Rappelons l'article 30 (droits d'exploitation secondaire) de la Loi numérique de 2016 qui fixe la durée d'embargo à six mois pour les sciences de la technique et de la médecine (STM) et à douze mois pour les sciences humaines et sociales (SHS).

11 HAL en 2018 : 10 dépôts à l'heure. En ligne : <https://www.ccsd.cnrs.fr/2019/02/hal-en-2018-10-depots-a-l-heure/>.

12 Rapport d'activité 2019 du CCSD. En ligne : [https://www.ccsd.cnrs.fr/bilan\\_2019/](https://www.ccsd.cnrs.fr/bilan_2019/).

13 Voir le portail CNRS : <https://cnrs.hal.science/browse/doctype>.

14 Il est par exemple impossible d'obtenir des chiffres fiables concernant la taille de tous ces laboratoires.

15 Cluster 1 : >300 dépôts, >40% texte intégral ; Cluster 2 : >700 dépôts, <20% texte intégral ; Cluster 3 : <100 dépôts (voir Schöpfel, 2020).

16 Test statistique khi-deux d'indépendance ( $p \leq 0,05$ ).

17 Le tableau reprend les dix domaines scientifiques du RNSR. En ligne : <https://appliweb.dgri.education.fr/rnsr/>.

18 Les domaines correspondent aux champs de formation universitaires : Arts, Lettres, Langues, Sciences humaines et sociales (ALLSHS) ; Droit, Économie, Gestion (DEG) ; Sciences de la vie et de la santé (SDV santé) ; Sciences et technologies (ST).

19 Voir le CCSD pour les projets en cours. En ligne : <https://www.ccsd.cnrs.fr/projets/>.

## RÉSUMÉS

---

## **Français**

L'article présente les résultats d'une étude menée dans le cadre du projet HAL/LO, sur un échantillon de 1 246 laboratoires (=1 035 612 dépôts) rattachés aux dix grandes universités de recherche et membres de l'association Udice. L'objectif est une description plus détaillée des pratiques sur HAL. 99 % des laboratoires sont présents sur HAL, avec une distribution du type « longue traîne ». 52 % des publications sont des articles, 23 % des communications. Le degré d'ouverture moyen est 32 % (dépôts avec documents). 50 % des laboratoires ont créé une collection sur HAL. La discussion porte sur trois aspects : le rôle des laboratoires par rapport à HAL, avec une description plus détaillée de plusieurs situations types ; l'impact des disciplines par rapport au nombre des dépôts, à la création d'une collection, au dépôt de certains types de documents ou à l'auto-archivage des documents en texte intégral ; l'évolution du dispositif HAL vers un outil pour recenser la production scientifique, ce qui pose plusieurs questions notamment sur la provenance et la qualité des métadonnées.

## **English**

The article presents the results of a study carried out as part of the HAL/LO project, on a sample of 1 246 laboratories (=1 035 612 deposits) affiliated to the ten major research universities and members of the Udice association. The aim is to provide a more detailed description of practices on HAL. 99 % of laboratories are present on HAL, with a "long tail" distribution. 52 % of publications are articles, 23 % communications. The average degree of openness is 32 % (deposits with documents). 50 % of laboratories have created a collection on HAL. The discussion focused on three aspects: the role of laboratories in relation to HAL, with a more detailed description of some typical situations; the impact of disciplines in relation to the number of deposits, the creation of a collection, the deposit of certain types of documents and the self-archiving of full-text documents; the evolution of the HAL system towards a tool for monitoring scientific production, which raises several questions, notably concerning the origin and quality of metadata.

## **INDEX**

---

### **Mots-clés**

Science ouverte, archives ouvertes, accès ouvert, laboratoires de recherche, HAL

### **Keywords**

Open science, open repositories, open access, research laboratories, HAL

## AUTEURS

---

**Joachim Schöpfel**

Maître de conférences, université de Lille, GERiiCO,

**Florence Thiault**

Maître de conférences, université de Rennes 2, PREFics et GERiiCO,

**Hélène Prost**

Ingénieur d'études, CNRS, GERiiCO,

**Bernard Jacquemin**

Maître de conférences, université de Lille, GERiiCO,

**Éric Kergosien**

Maître de conférences, université de Lille, GERiiCO