



Augmenter la découvrabilité scientifique : accompagner l'auto-archivage dans HAL ?

Alaric Tabariès

► To cite this version:

Alaric Tabariès. Augmenter la découvrabilité scientifique : accompagner l'auto-archivage dans HAL ?. Sciences de l'information et de la communication. Université de Toulon, 2025. Français. <NNT : 2025TOUL4004>. <tel-05587857>

HAL Id: tel-05587857

<https://theses.hal.science/tel-05587857v1>

Submitted on 10 Apr 2026

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.



HAL Authorization

ÉCOLE DOCTORALE N°509
SOCIÉTÉS MÉDITERRANÉENNES ET SCIENCES HUMAINES
Institut Méditerranéen des Sciences de l'Information et de la
Communication (EA 7492)

THÈSE présentée par :

Alaric TABARIES

soutenue le : 13 novembre 2025

pour obtenir le grade de Docteur en Sciences de l'Information et de la
Communication

**Augmenter la découvrabilité scientifique :
accompagner l'auto-archivage dans HAL ?**

THÈSE dirigée par :

REYMOND David

Maître de conférences HDR, Université de Toulon

Co-encadrée par :

CUXAC Pascal

Ingénieur de recherche, INIST CNRS

JURY :

ROUISSI Soufiane

Professeur des universités, Université Bordeaux
Montaigne

SZONIECKY Samuel

Maître de conférences HDR, Université Paris 8

SCHAFFAR-DIMOU Alexandra

Professeure des universités, Université de Toulon

VERLAET Lise

Professeure des universités, Université Paul-
Valéry Montpellier 3

Résumé

Cette thèse interroge les pratiques d'auto-archivage scientifique dans le contexte français, à travers une analyse approfondie de l'archive ouverte HAL. Elle articule une approche socio-technique et informationnelle pour examiner les dynamiques de dépôt, la qualité des métadonnées et les logiques de découvrabilité associées. S'appuyant sur une recherche-action menée à l'Université de Toulon et sur le développement du dispositif SoVisu, ce travail met en lumière les tensions entre prescriptions institutionnelles et pratiques effectives, ainsi que les spécificités disciplinaires, notamment en sciences de l'information et de la communication. Il propose une typologie des contributeurs et des logiques de dépôt, tout en explorant les enjeux de visibilité, de traçabilité et d'évaluation dans le cadre de la science ouverte. En croisant analyse quantitative massive et réflexivité méthodologique, cette recherche contribue à une meilleure compréhension des conditions d'appropriation des archives ouvertes et de leur rôle dans l'écosystème de la communication scientifique contemporaine.

Mots-clés : auto-archivage ; archives ouvertes ; découvrabilité ; science ouverte ; HAL ; métadonnées ; visualisation ; communication scientifique ; information scientifique et technique ; pratiques informationnelles

Abstract

This thesis investigates self-archiving practices within the French academic context by focusing on the open archive HAL. Adopting a socio-technical and informational approach, it explores deposit dynamics, metadata quality, and the underlying logics of discoverability. Grounded in an action research project conducted at the University of Toulon and centered on the development of the SoVisu platform, this work highlights the tensions between institutional guidelines and actual practices, with particular attention to disciplinary specificities in information and communication sciences. It proposes a typology of contributors and deposit logics, while examining issues of visibility, traceability, and evaluation within the broader framework of open science. By combining large-scale quantitative analysis and methodological reflexivity, this research provides a deeper understanding of how open archives are appropriated and their role in shaping contemporary scholarly communication.

Keywords : self-archiving ; open archives ; discoverability ; open science ; HAL ; metadata ; visualization ; scholarly communication ; scientific and technical information ; information practices

Remerciements

Mes tous premiers remerciements s'adressent à mon directeur de recherche, David Reymond qui m'a accompagné dans cette aventure qu'est la thèse de doctorat, et sans qui, je n'aurais jamais découvert le monde, au combien passionnant, de la recherche universitaire. Je tiens à exprimer ma profonde gratitude à David pour sa guidance avisée, sa bienveillance et sa gentillesse au cours de ces années. Je remercie au même titre Pascal Cuxac, pour son regard rigoureux, ses conseils éclairés et sa disponibilité constante.

Je tiens à mentionner les chercheuses Annaïg Mahé et Camille Prime-Claverie qui, bien que nous n'ayons pas eu d'échanges directs, ont contribué à orienter ma réflexion, à un moment où l'expérimentation appelait à des repères théoriques et méthodologiques plus solides.

Je remercie également mes collègues du laboratoire IMSIC et de l'UFR Ingémédia pour les riches échanges que nous avons pu avoir, tant sur le plan académique, qu'humain. À ce titre, je tiens à remercier plus particulièrement Sami Ben Amor pour m'avoir accordé sa confiance dans le cadre de mes deux années en tant qu'attaché temporaire d'enseignement et de recherche. Je n'oublie pas Michel Durampart qui a toujours fait preuve d'une bienveillance et offert des conseils avisés lors de nos échanges.

Mes pensées vont également à mes parents qui m'ont accompagné tout au long de mes études. À Vinciane, ma compagne, qui a accepté mes soirées d'écriture devant la douce lumière bleue de l'écran d'ordinateur.

Enfin, je remercie également les membres de ce jury qui ont accepté d'évaluer ce travail.

Sommaire

Introduction.....	1
Positionnement et choix rédactionnels.....	2
Partie 1 - Histoire, développements et impacts du libre accès.....	5
1.1. Des échanges épistolaires à la crise des périodiques.....	6
1.2. Le libre accès à l'information scientifique.....	8
1.3. Les archives ouvertes.....	44
Partie 2 - SoVisu, un dispositif d'accompagnement à l'auto-archivage.....	53
2.1. Contexte.....	54
2.2. Un premier prototype fonctionnel.....	55
2.3. Extension du développement dans une logique d'accompagnement.....	59
2.4. Limites.....	79
2.5. Conclusion.....	80
Partie 3 - Cartographie des pratiques d'archivage sur HAL.....	83
3.1. Méthodologie.....	84
3.2. Résultats.....	99
3.3. Discussion.....	145
Conclusions et perspectives.....	149
Bibliographie.....	156

Table des figures

Figure 1.	Les grandes étapes du développement du libre accès	35
Figure 2.	Cycle du processus de recherche	41
Figure 3.	La taxinomie de la science ouverte	43
Figure 4.	« Les temps forts de la vie de HAL », première partie	47
Figure 5.	« Les temps forts de la vie de HAL », seconde partie	48
Figure 6.	Capture d'écran du tableau de bord du premier prototype	56
Figure 7.	Capture d'écran de la liste des publications catégorisées du premier prototype	56
Figure 8.	Capture d'écran de l'exploration du réseau de collaboration du premier prototype	57
Figure 9.	Capture d'écran de l'exploration terminologique du premier prototype	58
Figure 10.	Schéma fonctionnel du dispositif SoVisu	68
Figure 11.	Capture d'écran de la page « vérifier ses données » du dispositif SoVisu	70
Figure 12.	Capture d'écran de la page « tableau de bord » du dispositif SoVisu	72
Figure 13.	Capture d'écran de la page « références » du dispositif SoVisu	73
Figure 14.	Capture d'écran de la page « expertise » du dispositif SoVisu	74
Figure 15.	Capture d'écran de la page « lexique extrait » du dispositif SoVisu	75
Figure 16.	Capture d'écran d'un résultat d'export Hcéres effectué depuis le dispositif SoVisu	77
Figure 17.	Capture d'écran de la page « recherche d'expert » du dispositif SoVisu	78
Figure 18.	Processus et infrastructure de collecte	95
Figure 19.	Capture d'écran d'une partie d'un tableau de bord Kibana	96
Figure 20.	Évolution de la quantité de dépôts dans le temps en STM	102
Figure 21.	Évolution du nombre de contributeurs uniques dans le temps en STM	102
Figure 22.	Évolution des types de contributeurs dans le temps en STM	103
Figure 23.	Évolution des logiques de dépôt dans le temps en STM	105
Figure 24.	Évolution des logiques de dépôt dans le temps en STM déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)	107
Figure 25.	Évolution des logiques de dépôt dans le temps en STM déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)	109
Figure 26.	Évolution des niveaux de description des dépôts dans le temps en STM	111
Figure 27.	Évolution des niveaux de description des dépôts dans le temps en STM déposées selon des logiques d'archivage et de communication scientifique directe (haut) et selon des logiques de référencement et de recensement (bas)	112
Figure 28.	Évolution des niveaux de description des dépôts dans le temps en STM déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)	114
Figure 29.	Évolution des niveaux de description des dépôts dans le temps en STM déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)	115
Figure 30.	Évolution de la quantité de dépôts dans le temps en SHS	119

Figure 31.	Évolution du nombre de contributeurs uniques dans le temps en SHS	119
Figure 32.	Évolution des types de contributeurs dans le temps en SHS	120
Figure 33.	Évolution des logiques de dépôt dans le temps en SHS	122
Figure 34.	Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)	124
Figure 35.	Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)	125
Figure 36.	Évolution des niveaux de description des dépôts dans le temps en SHS	127
Figure 37.	Évolution des niveaux de description des dépôts dans le temps en SHS déposées selon des logiques d'archivage et de communication scientifique directe (haut) et selon des logiques de référencement et de recensement (bas)	129
Figure 38.	Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)	130
Figure 39.	Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)	132
Figure 40.	Évolution de la proportion de dépôts en SIC dans le corpus de données SHS, dans le temps	136
Figure 41.	Évolution de la proportion de contributeurs en SIC dans le corpus de données SHS, dans le temps	136
Figure 42.	Évolution des types de contributeurs dans le temps en SIC (haut) et en SHS hors SIC (bas)	138
Figure 43.	Évolution des logiques de dépôt dans le temps en SIC (haut) et en SHS hors SIC (bas)	140
Figure 44.	Évolution des niveaux de description des dépôts dans le temps en SIC (haut) et en SHS hors SIC (bas)	141
Figure 45.	Évolution des notices décrites par à la fois un résumé et des mots-clés (haut) et celles non décrites (bas) depuis 2021	144

Introduction

Les transformations de la communication scientifique se déploient dans un contexte marqué par des injonctions parfois paradoxales : rendre les savoirs accessibles à tous, garantir leur fiabilité face à la prolifération des fausses informations, et renforcer la transparence des pratiques de recherche au nom d'un « juste retour » vers la société. Dans ce paysage en recomposition, les archives ouvertes occupent une position stratégique. À la fois instruments de diffusion directe et vecteurs de visibilité indirecte, notamment en étant moissonnées par des systèmes automatisés tels que les moteurs de recherche aujourd'hui, les systèmes d'intelligence artificielle, qui les exploitent notamment pour entraîner des modèles, elles jouent un rôle central dans la circulation des productions scientifiques.

Cette recherche naît d'une expérience concrète : le développement d'un prototype de visualisation, SoVisu, réalisé initialement dans le cadre d'un stage de Master, au sein du laboratoire IMSIC. Ce projet vise à outiller les services de soutien à la recherche pour mieux rendre visible l'information scientifique qui se retrouve sur l'archive ouverte HAL. C'est en concevant et en manipulant cet outil, dans une approche d'abord technique et pragmatique, que nous commençons à formuler un ensemble de questionnements plus théoriques et critiques : que déposent les chercheurs ? Selon quelles logiques ? Comment la complétude des métadonnées influe-t-elle sur la découvrabilité ? Et de quelle manière peut-on accompagner ces pratiques sans les normer de façon prescriptive ?

La problématique s'élabore ainsi progressivement, à la croisée de plusieurs tensions : entre prescriptions institutionnelles et usages concrets, entre dispositifs techniques et contextes d'appropriation, entre injonctions à la visibilité et réalités de contribution. En nous appuyant sur une posture de recherche-action et une démarche réflexive, nous cherchons à comprendre comment les chercheurs s'emparent, ou non, de HAL, et comment rendre visibles ces usages pour mieux les accompagner dans une logique d'ouverture maîtrisée.

Sur le plan méthodologique, ce travail articule une phase de développement technique, une analyse exploratoire quantitative à large échelle, et mobilise des investigations qualitatives ciblées. Le dispositif SoVisu, au cœur de la première phase de la recherche, constitue à la fois un outil, un objet d'étude et un déclencheur de problématisation. L'ensemble s'inscrit dans une approche socio-technique propre aux sciences de l'information et de la communication, attentive aux interactions entre normes, outils, pratiques et acteurs.

Cette thèse propose plusieurs contributions : une typologie des logiques de dépôt, un indicateur de découvrabilité, une cartographie des profils de contributeurs, et un dispositif interactif de médiation à visée pédagogique et réflexive. Elle vise à outiller la compréhension des usages réels de l'archive HAL, en éclairant les conditions concrètes de circulation, de description et de valorisation des savoirs scientifiques.

Ce manuscrit ne s'inscrit pas dans une logique hypothético-déductive classique, mais repose sur un ancrage empirique et progressif. Sa structuration en trois parties reflète une dynamique exploratoire, où la démarche de recherche se construit au fil des situations rencontrées, des outils mobilisés et des données analysées. Il ne s'agit pas de vérifier une théorie préalablement formulée, mais de faire émerger des questionnements à partir du terrain, dans une posture de recherche-action assumée.

La première partie offre un cadrage théorique et contextuel sur le libre accès et la science ouverte, permettant de situer historiquement et politiquement l'émergence de dispositifs comme HAL. Ce détour est nécessaire pour comprendre les enjeux documentaires et informationnels en jeu. La deuxième partie s'inscrit dans la continuité directe de mon engagement de terrain, en restituant la genèse et les évolutions du dispositif SoVisu, à la fois objet technique, support d'analyse et outil de médiation. Cette partie ne vise pas seulement à décrire un outil, mais à interroger ce que sa conception révèle des pratiques, tensions et besoins autour de l'auto-archivage. Enfin, la troisième partie propose une montée en généralité à partir d'une analyse à large échelle des dépôts HAL, avec une attention particulière aux effets d'un accompagnement ciblé sur les logiques de contribution et la qualité des métadonnées. Chaque partie est ainsi articulée à la précédente dans un enchaînement qui cherche à conjuguer immersion locale, conceptualisation et extrapolation méthodique.

Positionnement et choix rédactionnels

Avant d'entrer plus en profondeur dans le sujet, il convient de préciser certains choix rédactionnels adoptés dans ce manuscrit. Certains passages sont rédigés à la première personne du singulier. Ce choix, encore peu fréquent dans les écrits académiques en sciences de l'information et de la communication, répond à une volonté de transparence méthodologique et de réflexivité. Il vise à expliciter certaines étapes du travail de recherche en adoptant un ton personnel, en particulier lorsqu'il s'agit de relater des expériences de terrain, des choix méthodologiques ou des enjeux rencontrés dans la conduite de l'enquête. Ces moments narratifs

permettent de rendre compte plus finement du cheminement du chercheur, en assumant une posture située. Néanmoins, conformément aux usages académiques, la majeure partie de cet ouvrage est rédigée à la première personne du pluriel, afin de conserver une posture analytique et distanciée. Ce passage du « je » au « nous » n'est pas anodin : il reflète une articulation entre une démarche personnelle engagée et l'inscription dans un cadre scientifique collectif. Par ailleurs, certains choix d'écriture ont été pensés pour faciliter la lisibilité du propos, notamment par l'introduction de transitions interprétatives au sein de l'analyse, par une structuration typologique des résultats et par l'articulation d'approches quantitatives et qualitatives. Il s'agit ici de rendre accessibles des phénomènes complexes sans en sacrifier la rigueur, en assumant une volonté de médiation scientifique.

Cette courte introduction personnelle vise à éclairer le regard que je porte sur cette recherche et à permettre d'en appréhender certains choix sous un prisme singulier. Ces années de thèse constituent l'occasion d'un travail réflexif nourri, non seulement sur les objets étudiés, mais également sur ma propre position en tant que chercheur engagé dans une démarche de recherche-action. Cette posture implique de considérer l'influence que peut avoir l'expérience personnelle sur la démarche de recherche. Ce travail réflexif se révèle essentiel pour penser ma place dans le processus de recherche, en cohérence avec l'attention portée, en sciences de l'information et de la communication, aux postures situées et aux conditions de production des savoirs. Issu d'un environnement où la curiosité intellectuelle était valorisée, je garde en mémoire les expérimentations de mon adolescence, comme le démontage d'objets techniques pour en comprendre le fonctionnement, prémices d'une posture exploratoire qui traverse aujourd'hui mon rapport à la recherche. À l'adolescence, cette curiosité se tourne vers l'outil informatique, et j'explore les possibilités infinies offertes par le développement. Cette inclination pour l'expérimentation concrète et l'exploration autonome explique qu'à plusieurs reprises, au fil de mes études supérieures, j'ai envisagé de rejoindre le monde professionnel : d'abord après l'obtention de mon diplôme universitaire de technologie en métiers du multimédia et de l'Internet, puis de ma licence professionnelle en métiers des réseaux informatiques et télécommunications, et enfin à l'issue du Master information et communication. Pourtant, ce sont des rencontres déterminantes avec certains enseignants, mais aussi des circonstances plus fortuites, qui m'ont conduit à prolonger ce cheminement. Ce manuscrit représente aujourd'hui l'aboutissement d'un parcours universitaire engagé, mais en aucun cas la fin d'une quête intellectuelle. Il constitue plutôt une étape dans un processus réflexif en constante évolution, porté par une curiosité toujours vivace, et nourri par l'articulation constante entre pratique et analyse.

C'est dans cette perspective, à la croisée de la recherche-action, de l'ingénierie documentaire et de la réflexion critique sur les outils de la science ouverte, que s'inscrit la première partie de ce travail. Elle propose de revenir sur l'histoire du libre accès, ses fondements politiques, ses déclinaisons institutionnelles et ses effets concrets sur les pratiques d'auto-archivage. Ce détour théorique et contextuel permet de mieux situer l'archive ouverte HAL dans le paysage scientifique français, et d'éclairer les tensions qui en traversent l'usage.

Partie 1 - Histoire, développements et impacts du libre accès

« Ce mouvement nécessite de prendre des décisions qui seront stratégiques à la fois pour l'avenir de la recherche, et pour l'avenir de la façon dont la recherche s'inscrit dans la société et répond aux enjeux du monde contemporain. Il nécessite également que chaque acteur, à la place où il se situe aujourd'hui, agisse concrètement. »

(Dacos, 2016)

Dans la première partie de ce chapitre, nous recensons les initiatives en faveur du libre accès à l'information scientifique, et parfois à l'information de manière générale. Avec presque dix milliards de résultats aujourd'hui, la requête Google « *open access* » démontre à la fois l'immense profondeur du sujet mais aussi la complexité à recenser et retranscrire de manière exhaustive les projets, lois, textes et autres initiatives qui composent le mouvement. Fecher et Friesike (2014) rappellent d'ailleurs que la notion de science ouverte est fondamentalement polysémique : selon eux, la science ouverte est un terme générique qui englobe une multitude d'hypothèses sur l'avenir de la création et de la diffusion des connaissances, et peut être comprise de multiples façons, selon les acteurs ou les perspectives adoptées. Nous essayons donc, au fil de ce chapitre, de circonscrire le contexte de nos travaux. Compte tenu du sujet de cette thèse, l'accent est mis sur les initiatives françaises. Certaines initiatives ne portant pas directement sur le libre accès mais dont les implications pour le mouvement nous semblent importantes sont mentionnées. Le lecteur peut ainsi apprécier une vue d'ensemble des efforts entrepris par la communauté scientifique et les décideurs politiques pour promouvoir et façonner le libre accès tel que nous l'utilisons aujourd'hui.

Après cette vue d'ensemble, nous nous concentrons sur les éléments spécifiques au libre accès abordés dans cette thèse. Nous évoquons les différentes étapes du processus de création scientifique, en mettant en lumière celles qui subissent des changements significatifs induits par l'adoption du libre accès. Enfin, objet central dans une composante de ce processus, à savoir la diffusion de l'information scientifique, l'archive ouverte HAL fait l'objet d'une discussion approfondie.

1.1. Des échanges épistolaires à la crise des périodiques

Avant d'aborder la genèse et le développement du mouvement pour l'accès libre à l'information scientifique, il me semble important de comprendre la situation dans laquelle il apparaît.

La diffusion du savoir, essentielle au progrès scientifique, évolue au fil des époques que la société traverse et se transforme en réponse aux innovations technologiques ou culturelles. Jean-Claude Guéron, dans son article intitulé *L'histoire de la forme revue au prisme de l'histoire de la « grande conversation scientifique »*, retrace l'évolution de cette diffusion suivant différents paradigmes historiques. Nous rappelons ici les principales étapes de cette évolution.

Les échanges épistolaires, considérés comme le premier moyen de communication scientifique dans l'histoire, connaissent un changement majeur avec l'avènement de l'imprimerie. La capacité à

reproduire des écrits en quantité importante « mécanise » ces échanges et accélère la diffusion scientifique. Elle introduit conjointement la notion de marchandisation qui deviendra ensuite inhérente à l'imprimerie. Avec cette innovation technologique, et dès le XVIIe siècle, la revue scientifique moderne se dessine. C'est un premier tournant dans la diffusion de l'information scientifique.

Au XIXe siècle, avec l'émergence d'une communauté scientifique plus dispersée géographiquement, la communication scientifique s'organise autour d'associations scientifiques, aussi dénommées sociétés savantes. Les sociétés savantes répondent alors au besoin de centraliser et de diffuser les connaissances à travers un réseau étendu de chercheurs. Cependant, ces sociétés ne sont fondamentalement pas des maisons d'édition et n'en maîtrisent pas le savoir-faire. Ce déficit d'expertise induit une inefficacité dans le processus éditorial, et, permet aux maisons d'édition commerciales de commencer à se développer.

La forte croissance et la spécialisation de la recherche scientifique à la sortie de la seconde guerre mondiale éprouvent le modèle des maisons d'édition commerciales. Elles s'établissent conséquemment comme des acteurs incontournables de la diffusion de l'information scientifique. Parallèlement, avec l'avènement des nouvelles méthodes de bibliométrie telles que le facteur d'impact ou le h-index, les revues éditées par ces éditeurs acquièrent une dimension centrale au sein de la communauté scientifique, ce qui renforce davantage leur position dominante.

Avec l'irruption du numérique, les publications se dématérialisent. Les maisons d'édition opèrent alors une transition rapide vers un nouveau modèle économique, passant de la vente de produits physiques à la commercialisation de l'accès aux contenus dématérialisés. La commercialisation de l'information scientifique s'oriente, de fait, vers des abonnements numériques. Avec ces abonnements, la vente de « bouquets d'abonnements » (en anglais, *big deals*), où les éditeurs regroupent plusieurs revues en une seule et même offre, se développe. Bien que cette pratique réduit le coût même d'un article, les institutions se voient contraintes d'acheter l'accès à des revues moins pertinentes pour obtenir celles qui sont cruciales pour leurs besoins. Cette pratique, couplée à une inflation significative des prix des abonnements aux revues scientifiques, entraîne une hausse notable des dépenses pour les bibliothèques universitaires qui restent contraintes.

Cette situation marque profondément la communauté scientifique et engendre ce que l'on appelle la « crise des périodiques » (en anglais, *serials crisis*). Elle s'avère être un des catalyseurs pour repenser les modèles de diffusion du savoir et conduit à une remise en question des paradigmes traditionnels de l'édition scientifique.

1.2. Le libre accès à l'information scientifique

1.2.1. *Initiatives pionnières et premières archives ouvertes*

Le mouvement pour le libre accès à l'information scientifique et technique, qui vise à rendre les résultats de la recherche librement accessibles en ligne, sans barrière financière ou légale, émerge au cours des années 1990 dans un contexte de forte tension économique entre bibliothèques et éditeurs, souvent désigné sous le nom de « crise des périodiques ». Parmi les initiatives pionnières, la création en 1991 de la plateforme arXiv par Paul Ginsparg constitue un jalon déterminant. Inspiré des échanges informels facilités par les réseaux de forums tels que Usenet, ce projet permet aux chercheurs, initialement dans le domaine de la physique, de diffuser rapidement et librement leurs prépublications (en anglais, *preprints*), sans passer par les canaux éditoriaux traditionnels. En exploitant les potentialités d'Internet, arXiv contribue à rationaliser les processus de communication scientifique, tout en favorisant une culture de l'ouverture et de la transparence. Progressivement, la plateforme s'ouvre à d'autres disciplines et devient un modèle de référence, inspirant la création d'archives ouvertes similaires dans de nombreux pays.

Dans un élan similaire, Stevan Harnad lance CogPrints en 1997, une archive comparable dédiée aux travaux en sciences cognitives. Cette plateforme innovante accueille aussi bien les prépublications que les versions évaluées par les pairs, marquant une étape importante dans la démocratisation de l'accès aux travaux scientifiques dans ce domaine. Harnad joue également un rôle précurseur dans la promotion du libre accès en publiant la « Subversive Proposal » (1994). Sa proposition remet en question le modèle traditionnel de publication scientifique en invitant les chercheurs à auto-archiver leurs articles scientifiques, revus par des pairs, sur des serveurs accessibles au public et déclenche un débat important sur la propriété intellectuelle et l'accès à la recherche (O'Donnell & Okerson, 1995). Il soutient alors que cette démarche maximise la visibilité et l'impact de leurs recherches.

La Coalition de l'édition savante et des ressources académiques (en anglais, *Scholarly Publishing and Academic Resources Coalition*, SPARC), créée en 1997, s'inscrit dans cette dynamique d'ouverture. Elle s'implique pour réformer le modèle de communication scientifique en place, en se concentrant non seulement sur la promotion de l'accès ouvert auprès des pouvoirs publics, mais aussi en soutenant des modèles de publication plus équitables et transparents.

Pendant ce temps, des efforts similaires se développent au-delà des cadres traditionnels occidentaux. La Scientific Electronic Library Online (SciELO) créée en 1998 par le Brésil, se

focalise sur l'Amérique latine et facilite la publication en libre accès de revues scientifiques produites dans cette région du monde. Cette initiative renforce la visibilité internationale de la recherche régionale, permettant une circulation plus fluide des connaissances scientifiques au-delà des contraintes géographiques et linguistiques. L'émergence de cette bibliothèque illustre la pertinence universelle de l'accès libre aux résultats de recherche, en particulier dans les régions en voie de développement.

1.2.2. *Développement d'infrastructures pour l'interopérabilité*

En réponse aux défis posés par la diversité des formats et des protocoles, le mouvement de l'Open Archives Initiative (OAI), lancé en 1999, s'inscrit dans la continuité de ces initiatives fondatrices et répond au besoin d'harmonisation dans le domaine naissant des archives ouvertes. L'OAI établit un ensemble de protocoles et de techniques dans le but de faciliter l'interopérabilité entre les différentes archives. Ces actions rendent non seulement les résultats de recherche plus accessibles et découvrables pour la communauté scientifique internationale mais participent également à l'essor des dépôts. Cette méthodologie est formalisée dans le protocole Open Archives Initiative Protocol for Harvesting Metadata (OAI-PMH) en 2001.

À mesure que l'accès libre gagne en popularité et en nécessité, Harnad et son équipe à l'université de Southampton développent également EPrints, un logiciel de gestion de dépôts en libre accès. Il facilite la création de dépôts institutionnels et s'impose rapidement comme un des outils les plus utilisés dans ce domaine.

En France, la plateforme revues.org, lancée en 1999 par une association, devenue par la suite le Centre pour l'édition électronique ouverte (Cléo), sous l'impulsion, notamment de Marin Dacos, vise à promouvoir la visibilité et l'accès aux revues académiques en sciences humaines et sociales en utilisant le numérique comme moyen de diffusion. La plateforme permet la publication en ligne de revues, offrant ainsi un accès libre aux articles et aux numéros complets. Elle contribue fortement à la diffusion de la recherche en sciences humaines et sociales sur Internet.

Participant davantage à l'essor des infrastructures pour le libre accès, l'archive ouverte PubMed Central est créée en 2000 par le National Library of Medicine des États-Unis d'Amérique et se destine aux domaines des sciences de la vie et de la médecine. Cette archive sert de dépôt pour les articles scientifiques financés par le gouvernement étatsunien. Toujours dans les mêmes domaines, BioMed Central est lancé la même année et est le premier éditeur à offrir un accès libre à des revues de recherche biomédicales revues par des pairs en adoptant un modèle économique basé sur les frais de traitement d'article (en anglais, *Article Processing Charges*, APC). Avec cette

nouvelle voie de publication, les coûts de publication sont couverts par les auteurs ou leurs institutions et permettent une diffusion libre et ouverte des travaux.

L'idée commune de faciliter l'accès aux résultats de recherche qui motive ces démarches est reprise par la lettre ouverte de la Public Library of Science (PLOS) publiée en 2001. Cette lettre, rédigée par Patrick Brown, un biochimiste de Stanford et Michael Eisen, un bioinformaticien de Berkeley, et avec le soutien du prix Nobel Harold Varmus, est signée par plus de 30 000 scientifiques. Les auteurs préconisent un autre système de publication. Ils appellent les journaux à diffuser librement les articles au plus tard six mois après leur publication et invitent la communauté scientifique à s'investir dans la création de revues suivant ce modèle. Plusieurs journaux de premier plan, représentatifs de divers champs disciplinaires, tels que les *Proceedings of the National Academy of Sciences*, le *British Medical Journal* ou encore *Bioinformatics*, adoptent les propositions formulées par la PLOS soulignant l'impact significatif de cette initiative dans le monde de la recherche. Varmus, Brown et Eisen (2001) défendent une vision radicale de l'accès au savoir :

Nous pensons que l'enregistrement permanent et archivistique de la recherche scientifique et des idées ne devrait ni appartenir ni être contrôlé par les éditeurs, mais devrait appartenir au public et être librement accessible via une bibliothèque publique internationale en ligne.
(trad. pers.)

Cette citation condense une revendication fondatrice du mouvement pour le libre accès : l'idée que les résultats de la recherche, financés pour une large part sur fonds publics, doivent relever du bien commun. Elle engage une critique frontale des modèles propriétaires et propose une conception politique de la diffusion scientifique, où l'accessibilité devient un enjeu de justice épistémique. Ce principe, formulé avec force dès 2001, continuera d'irriguer les discours et les politiques en faveur de la science ouverte tout au long des deux décennies suivantes.

Dans la continuité des efforts pour faciliter l'accès à la recherche, les licences Creative Commons (CC), proposées dès 2002, fournissent des outils pour gérer les droits d'auteur de manière flexible. Ces licences permettent aux chercheurs de partager leurs travaux plus librement, en choisissant les conditions de réutilisation de leur production intellectuelle, ce qui accélère significativement l'adoption de pratiques de partage ouvertes parmi les communautés académiques. Elles sont notamment appliquées aux articles publiés par le PLOS.

Toujours dans le même prolongement, l'outil SHERPA RoMEO (*Rights Metadata for Open archiving*) est mis à disposition de la communauté. Cette ressource regroupe les politiques de libre

accès des revues et des éditeurs et en résume les autorisations d'auto-archivage et les conditions des droits accordés aux auteurs. En fournissant ces informations claires sur les droits de post-publication à la communauté scientifique, l'outil favorise une diffusion des travaux académiques.

La charte ECHO (en anglais, *European Cultural Heritage Online*), publiée en 2002, souligne l'importance de la préservation du patrimoine culturel et scientifique européen en s'appuyant sur les nouvelles technologies de l'information et de la communication qui facilitent un accès plus large à la connaissance. Elle encourage l'interopérabilité et la coopération dans le domaine de l'archivage ouvert.

En France, ces idées nouvelles se matérialisent avec la création du Centre pour la Communication Scientifique Directe (CCSD), sous l'impulsion de Franck Laloë, au sein du Centre national de recherche scientifique (CNRS) en 2000. La mission principale du CCSD est de « mettre à disposition des chercheurs de toutes disciplines, dans le monde entier, un circuit de communication directe leur permettant à la fois de gérer leurs propres résultats de recherche et de consulter ceux des autres chercheurs ». Le CCSD s'inspire du projet arXiv et crée un premier serveur Thèses en Ligne (TEL), basé sur le logiciel EPrints, en 2000. Ce serveur, dont le nom reflète directement sa fonction, est dédié à l'archivage des thèses de doctorat et des habilitations à diriger des recherches (HDR). Poursuivant ces développements, le CCSD met à disposition des chercheurs l'archive multidisciplinaire et institutionnelle Hyper Articles en Ligne (HAL) en 2001, en référence au supercalculateur éponyme dans le film *L'Odyssée de l'espace* de Kubrick, cette fois-ci sur un logiciel conçu en interne. La fonction principale de HAL est de promouvoir l'accès libre et ouvert aux résultats de recherche, en permettant aux chercheurs d'auto-archiver et de consulter des documents scientifiques de toutes disciplines. HAL est conforme au protocole OAI-PMH, ce qui signifie que ses métadonnées peuvent être moissonnées automatiquement par d'autres archives ou moteurs de recherche scientifiques. Cette conformité garantit l'intégration de HAL dans un écosystème global de circulation des savoirs, en facilitant l'indexation, la visibilité et la réutilisation des contenus déposés. Ainsi, certains dépôts, en fonction de leur domaine et de leur structuration, peuvent être automatiquement re-diffusés vers des plateformes comme arXiv, assurant une meilleure diffusion des travaux sans nécessiter de ressaisie. Comme le souligne Franck Laloë, initiateur du projet, « gérer des archives ouvertes, ce n'est pas qu'un problème technique [...], c'est avant tout une tâche très similaire à celle d'un éditeur d'une revue » (cité dans Veissier, 2019). Cette remarque rappelle que la mise en place de HAL engage des choix documentaires et politiques forts, dépassant la simple dimension logistique. Elle permet

également de comprendre certaines réticences initiales : certains chercheurs voyaient dans ce projet une menace de centralisation, allant jusqu'à comparer HAL à un outil de surveillance, par allusion ironique au film de Kubrick. Laloë raconte ainsi que l'un d'eux lui lança : « Ton projet est pire que HAL de l'Odyssée de l'espace » (ibid.).

Ces réticences ne se limitent pas à des peurs isolées. Dans certaines disciplines, elles ont conduit à des formes concrètes de prise d'autonomie vis-à-vis de HAL. C'est le cas notamment dans le domaine des sciences de l'information et de la communication des sciences de l'information et de la communication (SIC), où Gallezot, Chartron et Noyer, soutenus par le CNRS, créent en 2002 une archive ouverte spécifique aux SIC intitulée @rchiveSIC. Conçue à partir du logiciel EPrints et fondée sur le principe de l'auto-archivage, cette archive visait à répondre aux besoins propres de la communauté SIC. Les chercheurs ayant contribué à sa mise en place expliquent les motivations ayant guidé leur initiative :

Pour donner aux articles la lisibilité qu'ils méritent, pour faciliter la recherche et l'accès au savoir contenu dans les articles, pour relever le défi de la lutte contre la « babélisation » de la discipline tout en conservant la diversité qui en fait sa richesse, nous avons proposé à une instance de représentation de la communauté « information-communication » française la création d'une Archive ouverte : @rchiveSIC.

(Gallezot et al., 2003)

Dès sa conception, @rchiveSIC est pensée comme une interface disciplinaire, garante d'un filtrage éditorial et d'une reconnaissance symbolique adaptée aux SIC. Les auteurs insistent sur le fait que HAL devait être envisagé comme une infrastructure de destination, non comme une porte d'entrée directe : l'archive thématique devait jouer un rôle de médiation, tant documentaire qu'épistémique. Cette posture souligne une réticence initiale à déléguer à une plateforme généraliste la gestion de la visibilité des travaux en SIC, et justifie la création d'un dispositif spécifique, à la fois outil de structuration et instrument de représentation pour la communauté.

Nous revenons ci-après, dans ce même chapitre, sur la création de cette archive ouverte plus en détails.

1.2.3. Formalisation internationale du libre accès

Alors que ces développements témoignent une prise de conscience de l'importance du libre accès au sein de la communauté scientifique, l'Initiative de Budapest pour l'Accès Ouvert (en anglais, *Budapest Open Access Initiative* - BOAI) lancée en 2002 par des figures de proue telles que Stevan

Harnad, Michael Eisen, et également Peter Suber, formalise les idées et les principes du mouvement pour le libre accès et vise à accélérer les efforts en ce sens. L'initiative met en avant deux stratégies clés, en partie mises en œuvre dans les projets évoqués précédemment. La première est le développement d'archives ouvertes, conformes aux standards de l'OAI, permettant aux chercheurs d'auto-archiver leur production. La seconde encourage la création et le soutien de revues scientifiques en accès libre qui s'appuient sur le droit d'auteur pour ouvrir leur contenu plutôt que de le restreindre. La BOAI invite à explorer des pistes de financement alternatives au modèle traditionnel pour ces revues, en se tournant vers des sources diversifiées telles que les financements gouvernementaux ou universitaires. La citation suivante, extraite du texte fondateur, met en avant un aspect clé de la démocratisation du savoir, à savoir l'accès universel et libre à la littérature scientifique, quelles que soient les contraintes financières ou géographiques. Elle définit précisément le concept d'accès libre, en insistant sur les possibilités étendues qu'il offre :

« Par « accès libre » à la littérature, nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. »

(Budapest Open Access Initiative, 2002)

Cette définition souligne les implications fondamentales de l'accès libre, notamment la suppression des barrières financières, légales et techniques qui entravent la diffusion des savoirs. En facilitant un accès sans restriction aux résultats de la recherche, elle renforce l'implication des chercheurs, des enseignants, des étudiants, des professionnels et, plus largement, de toute personne souhaitant accéder, utiliser ou contribuer à la production scientifique et intellectuelle.

La même année, SPARC étend son rayon d'action en établissant SPARC Europe et œuvre davantage sur le plan académique européen.

1.2.4. Multiplication des déclarations et positionnements institutionnels

Dans la continuité de la BOAI, considérée comme le premier texte fondateur du mouvement, plusieurs déclarations se succèdent en 2003 pour renforcer son impact et encourager une adoption plus large de l'accès libre. Elles visent à répondre aux défis persistants liés à la diffusion des savoirs, notamment le coût croissant des abonnements aux revues scientifiques et les

restrictions d'accès imposées par les modèles traditionnels de publication. En incitant les institutions, les gouvernements et les chercheurs à adopter des pratiques favorisant le libre accès, ces déclarations contribuent à structurer et accélérer la transition vers une science plus ouverte et accessible à tous.

Clarifiant davantage les attentes autour de la publication en libre accès, la déclaration de Bethesda pour l'édition en libre accès précise clairement ce qu'implique une publication en libre accès. Selon cette définition, une publication en libre accès doit répondre à deux critères principaux. Les auteurs accordent un droit d'accès gratuit à tous, irrévocable, mondial et perpétuel, ainsi qu'une licence pour, entre autres, copier, transmettre et utiliser publiquement l'œuvre à condition d'être cités. La publication doit également être déposée sur une archive en ligne, soutenue par une institution académique, une société savante, un organisme gouvernemental, ou une autre organisation reconnue, qui utilise les standards techniques assurant l'accès libre à long terme.

La déclaration de Berlin sur le libre accès à la connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales, rédigée à la suite d'un congrès organisé par la société Max-Planck, confirme la définition d'une publication en libre accès proposée par la déclaration de Bethesda et l'étend à tout type de contribution scientifique, c'est-à-dire, à l'ensemble des résultats de recherche. Cette déclaration est signée par un nombre conséquent d'universités à l'échelle internationale. Elle affirme notamment que :

Notre mission de diffusion du savoir n'est remplie qu'à moitié si l'information n'est pas rendue largement et aisément accessible à la société.

(Max-Planck-Gesellschaft, 2003, trad. pers.)

Cette phrase emblématique résume l'esprit du mouvement : l'accès libre n'est pas un simple supplément technique ou documentaire, mais une condition de légitimité de la diffusion scientifique. Elle traduit un déplacement éthique du rôle des institutions de recherche, qui ne doivent pas seulement produire du savoir, mais aussi garantir sa circulation effective dans l'espace public.

La déclaration de principe du Sommet Mondial sur la Société de l'Information (SMSI) souligne l'importance cruciale de l'accès universel à l'information et au savoir pour bâtir des sociétés inclusives et démocratiques. Ce sommet vise à réduire la fracture numérique mondiale et à poser les bases d'une gouvernance internationale des technologies de l'information, tout en cristallisant des tensions entre une vision descendante, orientée vers la diffusion massive de l'information, et des approches plus ascendantes, portées par des acteurs associatifs et académiques, défendant une

conception de la société fondée sur la production partagée de connaissances. Bien qu'elle porte sur l'ensemble des formes d'information, la déclaration finale met ainsi l'accent sur des principes qui renforcent l'idée que le libre accès aux résultats de la recherche scientifique constitue un levier central pour le partage des savoirs, la promotion de l'innovation et le développement durable.

Il est particulièrement éclairant de revenir ici sur les tensions épistémologiques et politiques qui ont traversé les débats lors du deuxième SMSI, tenu à Tunis en 2005. Ce sommet, censé promouvoir une société de l'information inclusive et équitable, a donné lieu à des confrontations vives entre deux conceptions profondément divergentes. D'un côté, une approche institutionnelle et descendante, portée par les États et les grandes organisations internationales, conçoit la société comme façonnée par la diffusion massive de l'information, approche largement héritée d'un paradigme occidental et libéral. De l'autre, des voix issues de la société civile, d'ONG, d'associations mais aussi de chercheurs, notamment du Sud global, mais aussi relayées par certains acteurs critiques du Nord, ont défendu une vision plus ascendante, fondée sur la capacité des individus et des communautés à produire leurs propres connaissances, et à construire collectivement une société de la connaissance. Cette opposition, bien que schématiquement formulée ici, ne se réduit pas à une lecture binaire Nord-Sud, même si ces tensions géopolitiques ont traversé les discussions. Comme le soulignent Ambrosi, Peugeot et Pimienta dans leur ouvrage *Enjeux de mots : regards multiculturels sur les sociétés de l'information* (Ambrosi et al., 2005), le SMSI a cristallisé des divergences profondes autour de la place des savoirs, de leur mode de production, et des conditions de leur circulation. Michel Durampart, dans *Société de la connaissance, fractures et évolutions* (2009), revient sur ces débats en soulignant que « ces conceptions ne pouvaient que susciter une critique salutaire venue non seulement des milieux scientifiques mais aussi des représentants de la société civile qui peinent à se faire entendre dans les débats publics. La notion de société de connaissance est donc un axe épistémologique fertile ». Cette controverse n'est pas sans résonance avec la posture que nous adoptons dans ce travail. En effet, à travers l'étude des pratiques d'auto-archivage et des dynamiques de découvrabilité associées, nous défendons l'idée que les connaissances ne circulent pas de manière neutre ou désincarnée. Elles sont d'abord produites par les chercheurs, mais leur visibilité, leur accessibilité et leur appropriation dépendent ensuite des circuits d'information dans lesquels elles s'inscrivent. C'est en ce sens que nous nous inscrivons dans une conception critique et située de la société de la connaissance, proche de celle qu'évoque André Gorz (2003), pour qui les connaissances désignent « des contenus formalisés, objectivés, qui ne peuvent, par définition, appartenir aux personnes », et qui nécessitent dès lors des dispositifs de médiation, de contextualisation et de réappropriation pour prendre sens au sein de l'espace public.

Nous poursuivons ainsi avec la déclaration de la Fédération internationale des bibliothécaires et d'institutions (IFLA) qui s'aligne sur cette vision en affirmant le rôle primordial des bibliothèques et des professionnels de l'information dans la facilitation de l'accès libre à l'information.

Toujours en 2003, l'Open Society Institute (OSI) lance plusieurs projets relatifs au libre accès, dont le Directory of Open Access Journals (DOAJ), une base de données bibliographiques qui recense les revues scientifiques en accès libre. Le DOAJ est créé pour augmenter la visibilité et la facilité d'utilisation des revues en libre accès par la communauté scientifique.

Dans le même temps, plusieurs organisations notables, comme le Wellcome Trust, une importante fondation caritative de recherche médicale britannique, ou encore le partenariat InterAcademy (en anglais, *InterAcademy Partnership*, IAP), un réseau mondial d'académies des sciences, émettent des conclusions similaires et prennent position en faveur du libre accès. Ces organisations affirment que l'accès ouvert est crucial pour la diffusion des connaissances scientifiques et insistent sur l'importance d'inclure les pays en développement dans ces initiatives. L'IAP souligne spécifiquement la nécessité d'adopter des stratégies inclusives qui favorisent l'équité et la diversité, reconnaissant que le libre accès joue un rôle essentiel dans la réduction des disparités à l'échelle mondiale.

Jusqu'ici, nous avons présenté comment le libre accès s'est structuré progressivement à travers différentes initiatives pionnières et une série de déclarations internationales. La suite du chapitre met en avant comment ces principes sont adoptés et mis en œuvre au niveau institutionnel et gouvernemental à travers le monde.

1.2.5. Expansion des politiques nationales et institutionnelles

Au-delà des acteurs institutionnels, les gouvernements des pays membres de l'organisation de coopération pour le développement économique (OCDE) s'engagent en faveur de l'accès ouvert aux données de recherche dans la déclaration sur l'accès aux données de la recherche financée par des fonds publics publiée en 2004. Les pays signataires, parmi lesquels les États-Unis, le Japon, l'Allemagne, le Royaume-Uni, la France et d'autres nations, de la déclaration reconnaissent notamment que l'accès ouvert à ces données « favorise le progrès scientifique ».

En parallèle de ces prises de position, la déclaration écossaise sur le libre accès de 2004 (en anglais, *Scottish Declaration on Open Access* ; Kidd, 2004) souligne l'engagement des institutions écossaises à promouvoir et à soutenir le mouvement du libre accès. Elle appelle à une transformation du modèle traditionnel de publication scientifique, en encourageant

l'auto-archivage et le soutien aux revues en libre accès et assure que ces transformations ne peuvent être que bénéfiques pour l'Écosse.

Les principes de Washington D.C. pour le libre accès à la science, énoncés en 2004 par des éditeurs scientifiques à but non lucratif, soulignent le rôle des éditeurs dans le mouvement. Ces éditeurs s'engagent à mettre en libre accès les travaux publiés soit immédiatement soit après un court délai suivant leur publication. Toujours en 2004, l'Association of Learned and Professional Society Publishers, composée aussi d'éditeurs à but non lucratif publie une position dans laquelle elle indique être favorable aux initiatives décrites précédemment telles que les revues en libre accès, les archives institutionnelles et l'auto-archivage) et appelle à étudier les retours en termes de bénéfices pour la communauté scientifique et de viabilité sur le long terme.

La même année, l'éditeur scientifique Springer, reconnaissant l'évolution rapide vers l'accès libre et répondant aux besoins croissants de la communauté scientifique, prend une initiative significative et devient le premier éditeur commercial à proposer un modèle de publication hybride. Ce modèle permet aux auteurs de choisir entre la publication traditionnelle par abonnement et la publication en libre accès moyennant le paiement d'APC. Cette ouverture stratégique reflète l'effort des grands éditeurs pour s'adapter sans rompre avec le modèle économique dominant :

Springer continue de proposer et recommande le modèle traditionnel de publication, éprouvé dans le temps pour garantir la qualité éditoriale et l'indépendance, mais offre désormais aux auteurs la possibilité de payer pour que leurs articles soient accessibles gratuitement à tous, partout dans le monde.

(« Springer goes head to head with Elsevier on open access », 2004, trad. pers.)

Loin de remettre en cause le système des abonnements, cette initiative propose une solution intermédiaire, où l'ouverture est conditionnée par la capacité de l'auteur, ou bien de son institution, à financer la diffusion immédiate de ses travaux. Elle illustre ainsi les tensions structurelles entre accessibilité, modèle économique et préservation du prestige éditorial. Plusieurs travaux ont analysé la complexité des dynamiques économiques et politiques associées au libre accès. Baruch (2007) met en lumière les tensions suscitées par les modèles hybrides expérimentés par certains éditeurs, notamment les dispositifs de paiement à la charge des auteurs, perçus comme une tentative d'adaptation commerciale à la science ouverte. Chartron (2010), pour sa part, éclaire la reconfiguration des équilibres entre acteurs institutionnels, chercheurs et opérateurs privés, révélant une redistribution du pouvoir dans l'espace documentaire académique.

L'engagement en faveur du libre accès s'affirme encore dans les pays en développement avec la déclaration de Salvador sur le libre accès adoptée en 2005 au Brésil. De par cette déclaration, les signataires demandent aux gouvernements de faire du libre accès une priorité de leur politique scientifique. Ils appellent les décideurs à rendre la recherche financée sur fonds public accessible en libre accès, à intégrer les coûts de publication (ou coûts de traitement d'article) dans les budgets de la recherche et à promouvoir les initiatives favorisant l'accès libre.

Le Wellcome Trust instaure dès 2005 une obligation de dépôt des travaux qu'elle finance dans les archives ouvertes Pubmed Central puis dans Europe PMC. La fondation développe en parallèle des mécanismes de financement qui permettent de couvrir les APC évitant ainsi d'imposer ces coûts aux auteurs.

En 2006, l'Appel de Riyadh (en arabe, *Nidae Al-Riyadh*) souligne aussi un engagement régional en faveur de la promotion du libre accès. Cette initiative vise à assurer un accès libre à la littérature scientifique, en particulier celle qui a une importance cruciale pour le passé, le présent, et l'avenir du monde arabe. Elle recommande à nouveau l'auto-archivage et l'utilisation de revues en libre accès comme moyens pour y parvenir.

L'OCDE poursuit son implication avec la publication des treize principes pour l'accès aux données de recherche financées par des fonds publics en 2007, l'engagement international en faveur du libre accès prend une nouvelle dimension. Les principes promulgués (je cite, « ouverture, flexibilité, transparence, conformité au droit, protection de la propriété intellectuelle, responsabilité formelle, professionnalisme, interopérabilité, qualité, sécurité, efficience, responsabilité de rendre compte et pérennité ») appellent à une plus grande ouverture et accessibilité des données de recherche, toujours dans l'objectif de favoriser l'innovation à travers une plus large collaboration scientifique dans le monde.

En 2008, le Japon, au travers du Japan's National Institute of Informatics (NII), publie le service web Japanese Institutional Repositories Online (JAIRO) et ouvre l'accès à son savoir académique. Le JAIRO permet de rechercher et d'accéder librement aux travaux scientifiques des différentes archives institutionnelles japonaises.

Aux États-Unis d'Amérique, une loi promulguée en 2009 rend obligatoire la mise en libre accès des publications issues des recherches financées par les National Institutes of Health (NIH). Les chercheurs doivent alors déposer les versions finales de leurs manuscrits sur PMC et le texte doit être accessible dans un délai de 12 mois suivant la publication.

Dans les traces de l'Amérique Latine et les pays arabes, l'Afrique s'inscrit dans le mouvement mondial du libre accès en mettant en place ses propres initiatives en faveur du libre accès. Une organisation sud-africaine à but non lucratif s'appuie sur sa plateforme African Journals OnLine (AJOL) créée en 1999, originellement destinée à promouvoir les revues africaines et diffuse dès 2009, les travaux signalés en accès libre.

En 2009, l'engagement de l'Union Européenne envers le libre accès entre dans une nouvelle dimension avec la mise en place de l'Open Access Infrastructure for Research in Europe (OpenAIRE). Cette initiative vise à renforcer le libre accès à travers l'Europe en facilitant la collaboration et le partage de ressources entre les pays membres. OpenAIRE fournit une infrastructure technique et des services de soutien pour aider les chercheurs à gérer et diffuser leurs travaux en libre accès, contribuant ainsi de manière significative à l'harmonisation des pratiques de recherche ouverte à l'échelle continentale. L'objectif est de maximiser l'impact et la visibilité des recherches financées par des fonds publics, en assurant un accès facile et ouvert aux résultats scientifiques pour tous les acteurs du domaine académique et au-delà.

La France est également un terreau fertile pour l'accès libre. Dans son rapport, le consortium Couperin, qui regroupe plus de 250 établissements de l'enseignement supérieur et de la recherche (ESR) intitulé *Open Access in France, a State of the Art Report* publié en 2010, référence 168 revues en libre accès et 68 archives thématiques ou institutionnelles.

La déclaration de Singapour sur l'intégrité en recherche, publiée en 2010, établit un cadre international pour la transparence dans la conduite et la communication de la recherche, l'honnêteté dans la publication des résultats, et la responsabilité dans l'utilisation des ressources de recherche. Bien que la déclaration ne traite pas directement du libre accès, elle énonce des principes qui ne peuvent qu'aller de pair avec un accès libre aux résultats de recherche.

L'initiative Riding the Wave - How Europe Can Gain from the Rising Tide of Scientific Data vient compléter ces principes en mettant l'accent sur l'importance cruciale de l'accès ouvert aux données scientifiques pour le progrès de la recherche. Publié en 2010 par le High Level Group on Scientific Data, ce rapport propose des orientations stratégiques pour la gestion et le partage des données scientifiques à travers le continent, soulignant la nécessité d'une infrastructure européenne dédiée au libre accès des données. Bien que développée indépendamment, l'infrastructure OpenAIRE matérialise ces propositions de par sa mission.

Les idées d'ouverture s'étendent à d'autres perspectives. Priem et al. (2010) proposent les *altmetrics*, ou métriques alternatives comme complément aux méthodes traditionnelles d'évaluation de

l'impact scientifique. Contrairement aux mesures basées uniquement sur le nombre de citations dans des revues académiques, les *altmetrics* prennent en compte une gamme plus large d'indicateurs reflétant l'engagement et l'influence d'une publication au-delà du milieu académique traditionnel. Ces indicateurs incluent, par exemple, les mentions sur les réseaux sociaux, les blogs, les discussions dans les médias grand public, ainsi que les téléchargements et les partages de documents en ligne. Les *altmetrics* offrent une perspective plus complète et nuancée de la manière dont les travaux scientifiques sont reçus et utilisés par différentes communautés, y compris les praticiens, les décideurs, et le grand public. En valorisant ces interactions, les *altmetrics* soutiennent l'idée que la recherche scientifique ne se limite pas à son évaluation par ses pairs, mais inclut également son influence dans le monde réel, contribuant ainsi à une évaluation plus juste et plus démocratique des contributions scientifiques.

En 2011, le paysage du libre accès en France s'enrichit avec le lancement d'OpenEdition, une plateforme numérique dédiée à la publication électronique dans le domaine des sciences humaines et sociales. OpenEdition est conçu pour offrir un accès ouvert à une large variété de contenus académiques, incluant des livres, des revues, des carnets de recherche, et des annonces scientifiques, consolidant ainsi le mouvement vers une diffusion plus accessible et transparente de la recherche. La plateforme se distingue par sa capacité à intégrer plusieurs formats de publications sous un même toit et par son engagement à promouvoir la visibilité des travaux en sciences humaines et sociales à l'échelle internationale.

Dans la dynamique de transformation des critères d'évaluation initiée avec les *altmetrics*, la Déclaration de San Francisco sur l'Évaluation de la Recherche (DORA), lancée en 2012, marque un tournant significatif dans la manière dont la communauté scientifique envisage l'évaluation de la recherche. Cette initiative, qui réunit un large éventail de parties prenantes, des chercheurs aux éditeurs et institutions académiques, vise à déplacer le focus des mesures traditionnelles de l'impact scientifique, telles que le facteur d'impact des journaux, vers une appréciation plus large et plus équitable de la valeur académique. DORA encourage la considération de diverses formes de contributions scientifiques et la valorisation de la qualité de la recherche plutôt que sa simple quantité ou le prestige du lieu de publication. En questionnant les pratiques d'évaluation existantes et en prônant des méthodes plus inclusives et transparentes, DORA soutient indirectement le mouvement pour le libre accès. En réduisant l'emprise des journaux à fort facteur d'impact, souvent payants, sur les publications, DORA ouvre la voie à une reconnaissance accrue des travaux publiés en libre accès, contribuant ainsi à une diffusion plus large et plus démocratique des connaissances scientifiques.

En 2012, le rapport Finch intitulé *Accessibility, sustainability, excellence: how to expand access to research publications*, commandé par le gouvernement britannique, recommande le développement d'une voie de publication durable garantissant à la fois la qualité et l'accessibilité des travaux. Les auteurs préconisent une transition vers le modèle basé sur les APC. Le rapport impacte significativement les politiques de recherche au Royaume-Uni et marque un tournant vers le soutien institutionnel à l'accès libre dans le pays.

Dix ans après la proclamation de l'Initiative de Budapest pour l'Accès Ouvert (BOAI), la communauté du libre accès se réunit à nouveau en 2012 pour évaluer les progrès accomplis et relever les nouveaux défis à travers les recommandations BOAI10. Les recommandations de la BOAI10 rappellent l'importance de renforcer les infrastructures d'archives ouvertes et de revues en accès libre, tout en encourageant la mise en œuvre de politiques de libre accès par les institutions de recherche, les financeurs et les gouvernements. Elle encourage à utiliser les licences ouvertes, telles que la licence CC-BY pour la publication, la distribution, la diffusion et la réutilisation des travaux scientifiques. Dans la continuité de la DORA, la BOAI10 amène aussi à penser des métriques alternatives pour évaluer l'impact et la qualité de la recherche.

Peter Suber, co-rédacteur de la BOAI, propose un état des lieux des développements de l'accès libre dans son livre *Open Access*. L'auteur offre une vue d'ensemble du mouvement. Il aborde les origines, les principes fondateurs, les défis rencontrés et les succès réalisés depuis les prémices du mouvement. Suber examine les différentes voies vers le libre accès, discute des modèles économiques afférents et des politiques institutionnelles et gouvernementales. Il mentionne la voie verte dans laquelle les chercheurs auto-archivent leurs travaux et la voie dorée où l'accès à l'information est gratuit pour le lecteur mais repose sur le paiement d'APC par les auteurs.

En 2013, la commission européenne imite les États-Unis d'Amérique avec l'article 29 du modèle de convention de subvention du programme Horizon 2020. La commission rend obligatoire la diffusion des résultats de recherche en accès libre financés par des fonds européens. Les chercheurs doivent déposer la version publiée ou la version acceptée par des pairs de leur publication dans une archive ouverte immédiatement et la rendre accessible librement dans les 6 mois pour les domaines des sciences, de la technique et de la médecine (STM) et 12 mois pour les sciences humaines et sociales (SHS).

La même année, le Royaume-Uni adopte la politique d'accès ouvert du Research Council UK (RCUK), un partenariat entre les sept conseils de recherche du pays. Cette politique oblige les chercheurs dont les travaux sont financés par le RCUK à publier leurs travaux en accès libre, soit

par la voie verte (6 mois après la publication pour les STM, 12 pour les SHS), soit par la voie dorée.

À l'échelle mondiale, les ministres chargés de la recherche, de la science, de la technologie et de l'innovation réunis au G8 Science soulignent l'importance de l'accès libre aux résultats de recherche financés par des fonds publics et la nécessité de les conserver de manière pérenne dans leur déclaration finale. Ils reconnaissent également les voies vertes et dorées pour y parvenir

Toujours sur le vieux continent, dans le cadre de sa collaboration avec OpenAIRE, l'organisation européenne pour la recherche nucléaire (CERN) lance la plateforme de dépôt de données de recherche Zenodo. Cette initiative vise à répondre aux besoins croissants de la communauté scientifique en matière de partage et de préservation des résultats de recherche. Elle offre un espace où les chercheurs de toutes disciplines peuvent librement déposer non seulement des publications scientifiques mais aussi des ensembles de données, des logiciels, et d'autres types de résultats de recherche, indépendamment du domaine scientifique.

En 2014, deux des plus influentes agences de financement chinoises adoptent également des politiques en faveur du libre accès. L'Académie chinoise des sciences demande à ce que les articles financés par des fonds publics soient accessibles librement dans les 12 mois après leur publication. La Fondation nationale des sciences naturelles de Chine applique cette même règle aux travaux qu'elle finance.

En 2015, le Canada emprunte le même chemin que les États-Unis, l'UE, le Royaume-Uni, la Chine, etc. avec la promulgation de la Tri-Agency Open Access Policy qui requièrent que les publications revues par les pairs et financées en partie ou totalement par les trois principaux organismes de financement public (les IRSC, le CRSNG et le CRS) soient diffusées en libre accès dans les 12 mois suivant leur publication.

D'autres pays adoptent de telles politiques au niveau national au début de cette décennie mais lister l'intégralité de ces initiatives serait excessivement long et hors du champ de cette thèse. Cependant, il est essentiel de reconnaître que l'adoption de politiques d'accès libre pour les travaux financés sur fonds publics par de nombreux pays souligne une tendance globale vers une plus grande transparence et accessibilité de la recherche scientifique.

À l'image de la DORA, le Manifeste de Leiden, publié en 2015, propose des principes pour guider l'évaluation de la recherche, en mettant l'accent sur la nécessité d'évaluer la recherche sur la base de son contenu propre plutôt que sur la base de métriques de publication telles que le

facteur d'impact des revues. . En recommandant des approches alternatives pour l'évaluation scientifique, le Manifeste de Leiden participe au mouvement pour le libre accès en promouvant une vision plus juste de la reconnaissance académique, qui valorise toutes les formes de contributions scientifiques, y compris celles rendues possibles grâce à l'accès libre.

En 2015, la Déclaration de La Haye sur l'extraction des connaissances marque une étape importante en faveur de l'exploitation des données numériques pour la recherche. Cette déclaration souligne la nécessité d'adopter des cadres juridiques et des politiques qui facilitent l'accès ouvert aux ensembles de données scientifiques pour permettre leur analyse et, en par extension, l'extraction de nouvelles connaissances.

Poursuivant l'effort de formalisation de l'accès aux données, les principes FAIR, formulés en 2016, proposent des principes directeurs pour gérer les données de recherche en vue de les rendre trouvables (*findable*), accessibles, interopérables et réutilisables par l'humain et les systèmes automatisés. Ils encouragent les acteurs de la recherche à adopter des pratiques qui facilitent un accès plus large et une utilisation plus efficace des données scientifiques dans le but de plus développer la collaboration scientifique et l'innovation.

La même année, le plan d'action d'Amsterdam sur l'innovation en matière de science ouverte (également appelé plan d'Amsterdam), représente un engagement fort de l'UE vers l'accès libre. L'appel recommande aux États membres d'adopter des politiques permettant l'accès libre à toutes les publications scientifiques ainsi que le partage et la réutilisation des données issues de recherches financées par des fonds publics.

Dans un effort de coordination globale pour réformer la communication scientifique, l'initiative internationale OA2020, initiée par la Max Planck Digital Library, marque un tournant dans l'adoption du libre accès à l'échelle mondiale. L'initiative souhaite rassembler les organismes de recherche pour transformer le système actuel de communication scientifique vers un accès ouvert total. En Allemagne, l'initiative OA2020-DE déclinée de cette même initiative est s'engage à promouvoir cette transformation au niveau national en privilégiant le conseil plutôt que des exigences strictes.

L'adoption de la loi pour une République numérique en 2016 constitue un tournant majeur pour la France dans sa politique de libre accès à l'information scientifique. Cette législation novatrice établit un cadre juridique favorable à la diffusion ouverte des résultats de la recherche financée par des fonds publics. Les auteurs peuvent désormais diffuser en libre accès les résultats de leurs

travaux de recherche financés à plus de 50 % par des fonds publics, après une période d'embargo de six mois pour les STM ou de douze mois pour les SHS. Comme le stipule l'article 30 :

Lorsque un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, est publié dans une revue paraissant au moins une fois par an, l'auteur dispose, même en cas de cession exclusive, du droit de le rendre disponible gratuitement sous forme numérique [...].

(Article 30 - LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1) - Légifrance, 2016)

Ce droit de mise à disposition publique renforce la légitimité juridique de l'auto-archivage et marque l'inscription formelle du libre accès dans le droit français.

Peu de temps après, Marin Dacos préface l'édition française du livre de Peter Suber *Open Access* et apporte un éclairage supplémentaire et contextualisé sur l'importance du libre accès, particulièrement dans le cadre européen et francophone. Il y rappelle que l'accès ouvert n'est pas une option technique, mais une transformation structurelle de la manière dont la recherche est produite, diffusée et gouvernée. Il insiste notamment sur la responsabilité de chaque acteur, affirmant que :

Ce mouvement nécessite de prendre des décisions qui seront stratégiques à la fois pour l'avenir de la recherche, et pour l'avenir de la façon dont la recherche s'inscrit dans la société et répond aux enjeux du monde contemporain. Il nécessite également que chaque acteur, à la place où il se situe aujourd'hui, agisse concrètement.

(Dacos, 2016)

Cette déclaration souligne que le libre accès engage une réorganisation des pratiques, au croisement de choix politiques, scientifiques et institutionnels. Elle renforce l'idée que le libre accès n'est pas seulement un cadre juridique ou un outil documentaire, mais un principe structurant des politiques scientifiques contemporaines.

En 2017, l'UE étend l'obligation de diffusion en libre accès pour les travaux de recherche financés par des fonds européens, émise en 2013, à l'ensemble des données de recherche dans le cadre du programme Horizon 2020. Cette nouvelle obligation s'intègre dans une démarche pour plus de transparence, de collaboration et d'innovation, en écho aux orientations définies, notamment, par le plan d'Amsterdam.

Dans le même temps, la déclaration de l'European Open Science Cloud (EOSC) établit un cadre pour la mise en place d'une infrastructure européenne de données de recherche, fédérée, conforme aux principes FAIR, visant à soutenir et à promouvoir l'accès libre à travers l'UE.

Dans les pays aux économies avancées, le G7 Science Ministers' Communiqué met en avant l'engagement en faveur de l'accès libre. Dans ce communiqué, les ministres chargés de la recherche, de la science, de la technologie et de l'innovation reconnaissent l'importance de l'accès libre aux données de recherche comme levier essentiel pour le progrès scientifique et l'innovation. Ils précisent que, pour atteindre cet objectif, le partage de ces données doit s'aligner sur les principes FAIR.

Cette même année, l'Appel de Jussieu pour la Science ouverte et la biodiversité marque une étape significative dans le débat sur l'accès libre et la diversité des publications scientifiques. Ce manifeste, soutenu par des chercheurs, des bibliothécaires et des éditeurs, souligne l'importance de développer des modèles économiques soutenables et équitables pour la publication scientifique, en opposition au modèle dominant des APC.

Face à une pression croissante pour des changements structurels dans la publication scientifique, cOAlition S, un consortium international d'organismes de financement de la recherche, lance le Plan S en 2018. Ce plan exige que, dès 2021, toutes les publications scientifiques résultant de recherches financées par des fonds publics soient publiées dans des revues ou des plateformes en libre accès, et ce, sans période d'embargo. Le plan stipule également que les chercheurs doivent retenir les droits d'auteur sur leurs travaux, assurant que les publications peuvent être librement partagées et réutilisées. Le Plan S vise à accélérer la transition vers un système de communication scientifique entièrement accessible et interopérable, mettant l'accent sur la nécessité pour la recherche financée par le secteur public d'être librement disponible sans barrières de paiement ou d'autres restrictions d'accès. Pour faciliter la mise en œuvre de ces exigences, de nombreux éditeurs et institutions académiques commencent à adopter des accords transformants, qui allouent les dépenses d'abonnement traditionnelles vers le financement de la publication en libre accès, alignant ainsi les pratiques éditoriales avec les principes du Plan S.

En France, le premier Plan National pour la Science Ouverte (PNSO) est lancé, développant davantage l'engagement de l'hexagone pour le libre accès. Le plan met en place des stratégies nationales pour promouvoir l'accès ouvert à la littérature scientifique et aux données de recherche. Ce cadre stratégique souligne la volonté de la France de participer à la généralisation de la science ouverte, alignée avec les politiques européennes et internationales, en mettant

l'accent sur la création d'un écosystème où la recherche financée par des fonds publics est largement accessible et réutilisable par toute la société. La citation suivante illustre le rôle central de l'archive ouverte HAL dans la stratégie nationale d'accès ouvert en France, en tant que référence incontournable pour l'auto-archivage des productions scientifiques, quelles que soient les disciplines concernées :

L'ouverture des publications scientifiques doit devenir la pratique par défaut aussi vite que possible. Pour engager cette dynamique, les publications issues de recherches financées au moyen d'appels à projets sur fonds publics seront obligatoirement mises à disposition en accès ouvert, que ce soit par la publication dans des revues ou ouvrages nativement en accès ouvert, soit par dépôt dans une archive ouverte publique comme HAL.

(Plan National pour la Science Ouverte, 2018)

Cette directive renforce l'ambition de généraliser l'accès libre aux résultats de la recherche publique, en établissant un cadre normatif pour garantir la diffusion et la pérennité des savoirs scientifiques. Confirmant cette dynamique, une enquête menée par le consortium Couperin (Ashta et al., 2020) révèle qu'en 2019, 23 % des établissements avaient instauré une obligation de signalement des références bibliographiques, et 15 % une obligation de dépôt du texte intégral. Ces taux marquent une progression significative par rapport à 2017, où seuls 9 % des institutions déclarent la première obligation, et 10 % la seconde. Ce rôle central qu'occupe HAL est d'autant plus singulier à l'échelle européenne que peu de pays ont fait le choix d'une telle centralisation technique et institutionnelle. Tandis que l'Allemagne, le Royaume-Uni ou les pays nordiques privilégient des plateformes décentralisées, souvent gérées au niveau des universités, la France a opté pour un dispositif unifié, adossé à une politique nationale de science ouverte. HAL constitue ainsi une exception notable, à la fois sur le plan technique et symbolique.

Le lancement de ce plan est d'autant plus pertinent que nos travaux démontrent que la loi pour une République Numérique contribue à augmenter le volume de dépôts sur l'archive nationale. Tabariès et Reymond (2023) soulignent ainsi que « la Loi pour la République numérique marque le début d'une phase exponentielle de contributions pour en dénombrer plus de 11 000 en SHS en 2020 ». Si cette dynamique est indéniable, elle ne saurait être interprétée comme un changement radical à elle seule. La croissance observée s'inscrit dans une évolution plus large, que nous analyserons plus loin, en tenant compte d'autres facteurs susceptibles d'avoir influencé les pratiques de dépôt sur le temps long. Les politiques gouvernementales successives jouent ainsi un rôle important en exerçant une pression continue et positive pour le dépôt des travaux de recherche en accès ouvert.

La même année, la France lance son premier baromètre de la science ouverte (Jeangirard, 2019). Cet outil de pilotage fournit des indicateurs détaillés sur l'ouverture des publications scientifiques, notamment le taux d'ouverture. Il permet ainsi d'ajuster les politiques en faveur de la science ouverte en mesurant leur impact. De notre point de vue, si ces métriques sont indispensables, elles ne permettent pas à elles seules d'assurer une évaluation complète des politiques d'ouverture. Nous en identifions les limites à la fin de ce chapitre et montrons l'utilisation de métriques complémentaires, afin de permettre une évaluation plus exhaustive.

Le rapport *Turning FAIR into reality*, publié la même année par la Commission européenne, recommande des actions concrètes pour l'adoption des principes FAIR, telles que le développement de politiques et de pratiques permettant une meilleure découverte des données de recherche, l'amélioration des infrastructures pour soutenir l'interopérabilité des données et la mise en place de services permettant la réutilisation efficace des données scientifiques. Il prône l'intégration de ces principes dès le début du cycle de vie de la recherche et souligne l'importance de l'éducation et de la formation pour sensibiliser les chercheurs à ces pratiques. Comme le précise le rapport :

Les chercheurs doivent être accompagnés par des infrastructures de données FAIR et des professionnels de la donnée, et ils doivent être formés et encouragés à gérer et partager leurs données selon les principes FAIR.

(European Commission, 2018, trad. pers.)

En France, ces recommandations se matérialisent notamment par la préconisation de rédiger des plans de gestion des données (PGD) pour les projets de recherche, inscrite dans le Plan national pour la science ouverte (PNSO).

En Europe également, la déclaration énoncée deux ans auparavant se matérialise avec le lancement de l'EOSC. Les chercheurs ont désormais accès à une infrastructure fédérée qui facilite la gestion, l'analyse et le stockage des données de recherche conformément aux principes FAIR.

En 2019, l'Initiative d'Helsinki sur le multilinguisme dans la communication scientifique marque un tournant important pour l'inclusion linguistique dans le libre accès. Cette initiative, soutenue par des organisations telles que l'UNESCO et le Conseil de l'Europe, vise à promouvoir l'égalité linguistique comme partie intégrante de la science ouverte. L'objectif est de permettre la publication de travaux scientifiques en plusieurs langues, tout en assurant leur disponibilité et visibilité à travers des archives ouvertes et des plateformes de libre accès. Cette démarche répond à un double enjeu : d'une part, elle favorise une plus grande diversité culturelle et linguistique

dans la recherche scientifique, reflétant ainsi une richesse de perspectives souvent absente dans les publications majoritairement anglophones ; d'autre part, elle renforce l'accès aux connaissances scientifiques pour les non-anglophones, facilitant ainsi une diffusion plus démocratique du savoir.

En 2019 également, le CNRS formalise son engagement envers la science ouverte à travers une feuille de route détaillée, visant à intégrer pleinement les principes de l'accès libre dans ses pratiques de recherche. Cette initiative s'inscrit dans une réponse aux enjeux scientifiques, sociaux et économiques, promouvant un accès ouvert aux résultats de recherche pour maximiser leur efficacité, visibilité et intégrité. Comme l'indique le document officiel :

La mise en œuvre de la feuille de route du CNRS pour la science ouverte [...] a l'ambition d'accélérer le processus vers la science ouverte en s'appuyant sur des actions concrètes structurées autour de quatre grands objectifs : garder le contrôle sur notre production scientifique et aboutir [...] à 100 % des publications du CNRS en accès ouvert ; développer une culture de la gestion/partage des données [...], développer des infrastructures [...] ; transformer l'évaluation individuelle des chercheurs [...]

(CNRS, 2019)

Le CNRS, en cohérence avec le Plan national pour la science ouverte lancé en 2018, établit ainsi des objectifs structurants autour de la publication, des données, des infrastructures et des modes d'évaluation.

1.2.6. Impact des crises sur l'accélération du libre accès

En 2020, la pandémie de COVID-19 accentue l'urgence d'un accès ouvert et immédiat à la recherche scientifique, soulignant l'importance vitale de la disponibilité rapide et transparente des résultats de recherche pour répondre aux crises sanitaires mondiales. Cette situation sans précédent met en lumière les limites des modèles traditionnels de publication scientifique et catalyse un mouvement mondial vers le libre accès, particulièrement dans les domaines liés à la santé et à la médecine. De nombreuses revues scientifiques, bases de données et consortiums de recherche réagissent en rendant leurs contenus relatifs au COVID-19 librement accessibles et en accélérant les processus de révision et de publication des études liées au virus. Cet élan vers un libre accès facilite une collaboration scientifique sans précédent et renforce les appels à un changement de paradigme dans la diffusion scientifique.

Suite à l'intérêt croissant pour le libre accès, poussé par la crise du COVID-19 en 2020, l'importance de la transparence des coûts associés à l'accès ouvert est devenue plus apparente. Dans ce contexte, plusieurs textes sont publiés et abordent ces enjeux financiers. Le projet *Open Access Price Transparency* cherche à clarifier et à standardiser les informations sur les coûts des publications en libre accès, y compris les APC. L'étude *APCs In The Wild* explore les différents flux de financement pour une transition accélérée vers l'accès ouvert, mettant en lumière la diversité et la complexité des modèles économiques en jeu. Enfin, l'état des lieux sur les accords transformants fournit une analyse détaillée de ces accords qui visent à convertir les abonnements traditionnels en modèles de financement entièrement ouverts. Ces documents et recherches conduisent à une meilleure compréhension des défis financiers de l'accès ouvert et formulent des stratégies pour rendre le libre accès plus viable et transparent pour la communauté scientifique.

Poursuivant cette réflexion sur les modèles économiques, le rapport *Read & Publish contracts in the context of a dynamic scholarly publishing system*, publié par l'association des universités européennes (en anglais, *European university association*, EUA), discute de l'essor des contrats Read & Publish et les problématiques afférentes, qui apparaissent en réponse à la pression croissante d'une partie des décideurs pour transitionner vers l'accès libre. Ces contrats sont des accords entre des institutions académiques et des éditeurs qui combinent les frais de lecture et de publication dans une seule et même transaction. Ces contrats permettent aux membres des institutions participantes d'accéder à l'ensemble des revues de l'éditeur (*Read*) et de publier leurs propres recherches en accès ouvert sans frais supplémentaires de traitement d'articles (*Publish*). L'objectif principal de ces contrats est de faciliter la transition vers le libre accès en rendant les recherches immédiatement disponibles à tous, tout en rationalisant les coûts associés à l'accès et à la publication scientifique. Comme le souligne le rapport de l'EUA :

Les accords Read & Publish peuvent soutenir la transition vers la publication en accès ouvert, mais ils ne représentent pas en eux-mêmes une transformation du système de publication, surtout s'ils se contentent de transférer les budgets d'abonnement vers la publication.

(European University Association, 2021, trad. pers.)

Le rapport met ainsi en garde contre le risque de reconduction de logiques anciennes, en particulier celles ayant mené à la crise des big deals.

La *Confederation of Open Access Repositories* (COAR) établi un cadre commun de bonnes pratiques pour les entrepôts, s'appuyant sur les travaux de l'Open Archives Initiative (OAI) pour standardiser et améliorer la gestion et la diffusion des données et publications scientifiques à

travers le monde. Ce cadre vise à renforcer l'interopérabilité entre les différents entrepôts en adoptant des protocoles ouverts et des standards développés par l'OAI, tout en promouvant des standards élevés en matière de métadonnées, d'accès aux données, et de préservation à long terme des ressources numériques. Les bonnes pratiques recommandées incluent l'implémentation de directives pour assurer la qualité et la fiabilité des dépôts, et le développement de politiques claires concernant les droits d'auteur et l'accès ouvert. L'objectif est de créer un réseau robuste d'entrepôts qui facilite l'accès universel à l'information scientifique et soutient les principes de la science ouverte, en diffusant les résultats de recherche selon les recommandations FAIR.

En France, le Comité d'Éthique du CNRS (COMETS) publie son avis intitulé « Les publications à l'heure de la science ouverte », traitant des implications éthiques et des défis associés à l'adoption de la science ouverte dans le contexte de la publication scientifique. Cet avis propose plusieurs recommandations pour harmoniser les pratiques de publication avec les principes de l'accès libre. Il met en avant la nécessité de trouver de nouvelles formes d'édition scientifique à l'image des épirevues, d'améliorer la transparence dans les processus de révision par les pairs, de soutenir l'archivage ouvert des publications et des données de recherche, et de modifier les critères d'évaluation des chercheurs pour valoriser les contributions à la science ouverte. En abordant ces points, le COMETS vise à promouvoir un système de publication qui non seulement accélère la diffusion des connaissances mais soutient aussi l'intégrité et l'équité dans la recherche scientifique.

L'étude de 2020, intitulée « Mesurer le taux d'accès ouvert des publications scientifiques : le cas de la France », présente une analyse détaillée de l'évolution du taux d'accès ouvert en France et offre un comparatif international. Cette recherche souligne les disparités importantes entre les différentes disciplines scientifiques et propose un indicateur d'accès ouvert normalisé qui tient compte de ces variations pour permettre des comparaisons équitables entre institutions ou pays, sans biais lié à la composition disciplinaire de leur production scientifique. Ce travail montre que bien que la France affiche un taux d'accès libre relativement élevé, il est important de normaliser ces mesures pour comprendre précisément la progression vers les objectifs de politique publique, notamment dans le contexte du Plan national pour la science ouverte qui vise un accès ouvert intégral à la production scientifique française.

1.2.7. Vers une généralisation de la science ouverte

En 2021, l'UNESCO adopte une recommandation notable sur la science ouverte, visant à rendre la science plus accessible, à renforcer la coopération scientifique mondiale et à améliorer la

confiance dans la science pour la société civile. Cette recommandation encourage les États membres à développer des politiques et des pratiques qui soutiennent les principes de la science ouverte. Elle préconise la promotion de l'accès ouvert aux publications scientifiques pour maximiser la disponibilité et l'impact de tous les résultats de recherche financés par des fonds publics. Parallèlement, elle met l'accent sur la facilitation de l'accès aux données de recherche afin d'accélérer les découvertes scientifiques et d'améliorer la transparence dans la recherche. L'UNESCO insiste également sur l'importance de développer des programmes d'éducation et de formation pour familiariser les chercheurs, les étudiants et le grand public avec les pratiques et les bénéfices de la science ouverte. Cette initiative représente un engagement global pour modifier la manière dont la science est pratiquée, partagée et perçue à travers le monde, alignant davantage les activités scientifiques avec les valeurs de transparence, d'inclusivité et de coopération.

The Principles of Open Scholarly Infrastructure (POSI), publié la même année, propose des directives essentielles pour le développement et la gestion des infrastructures de la recherche ouverte, visant à garantir que celles-ci soient durablement ouvertes. Les directives formulent des préconisations sur la gestion, le financement, l'ouverture et la protection des données à destination des organismes de financement, créateurs d'infrastructures et institutions de recherche pour promouvoir une infrastructure de recherche robuste et accessible.

Au niveau européen, deux rapports publiés par la Commission Européenne abordent la nécessité de réformer le système d'évaluation de la recherche et de faciliter l'engagement sociétal dans la science ouverte. Le premier rapport, *Towards a Reform of the Research Assessment System*, met en lumière les limites du système actuel, souvent centré sur des indicateurs quantitatifs comme les facteurs d'impact des publications, qui peuvent ne pas refléter fidèlement la qualité ou l'impact social de la recherche. Ce rapport appelle à une évaluation plus holistique qui reconnaît diverses formes de contributions scientifiques et qui encourage les pratiques alignées avec les principes de la science ouverte, telles que la collaboration, le partage de données et la publication en libre accès. Le second rapport, *Enabling Open Science and Societal Engagement in Research*, se concentre sur les stratégies pour intégrer l'engagement sociétal dans la recherche scientifique. Ce document souligne l'importance de la science ouverte non seulement pour la transparence et l'accessibilité, mais aussi pour son potentiel à impliquer activement les citoyens et les parties prenantes non académiques dans le processus de recherche. Il propose des recommandations pour développer des infrastructures, des politiques et des incitations qui soutiennent l'engagement du public et la co-crédation de connaissances, renforçant ainsi le lien entre science et société. Ces rapports reflètent une prise de conscience croissante de la nécessité d'adopter des modèles de recherche

plus inclusifs et transparents, soulignant un changement vers un paradigme où la science est plus accessible et davantage connectée aux besoins sociétaux.

En France, le deuxième Plan National pour la Science Ouverte (PNSO2) aligne ainsi la France avec les initiatives internationales. Le plan consolide l'engagement de l'hexagone envers la science ouverte en élargissant le libre accès non seulement aux publications financées totalement ou en partie par des fonds publics mais aussi aux données de recherche et aux codes sources. Le plan souligne également l'importance de cette ouverture pour les travaux issus des sciences humaines et sociales et renforce les moyens pour y parvenir. Il réaffirme aussi le rôle central de l'archive ouverte HAL dans la stratégie nationale, soulignant son importance pour la diffusion des résultats de recherche. De plus, le PNSO2 intègre les principes de la science ouverte dans plusieurs facettes de l'écosystème de la recherche, en particulier dans l'évaluation, le financement, et la publication. En révisant les référentiels du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres), le plan promeut des pratiques qui mettent en avant la qualité et l'impact sociétal des recherches, déplaçant ainsi l'accent mis sur le volume de publications ou leur prestige. Ces changements visent à renforcer la transparence et la collaboration au sein de la communauté scientifique française, tout en assurant une évaluation plus équitable des contributions des chercheurs. Comme le précise le document officiel :

Ce deuxième plan national étend son périmètre aux codes sources issus de la recherche, [...] multiplie les leviers de transformation afin de généraliser les pratiques de science ouverte [...] Il s'agit d'engager un processus de transformation durable afin de faire de la science ouverte la pratique commune et partagée.

(Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2021)

Cette déclaration reflète l'ambition systémique du PNSO2 : la science ouverte n'est plus envisagée comme un objectif périphérique, mais comme une norme intégrée à tous les niveaux de l'activité scientifique, appuyée par une volonté de transformation durable des pratiques et des politiques.

Poursuivant ces mêmes objectifs, l'Appel de Paris sur l'évaluation de la recherche, lancé en 2022, vise à réformer les pratiques d'évaluation académique à l'échelle mondiale. Cet appel se concentre sur la nécessité de repenser les méthodologies d'évaluation pour mieux aligner les critères avec les principes de la science ouverte. L'objectif est de promouvoir des systèmes d'évaluation qui valorisent la qualité et l'impact sociétal de la recherche plutôt que de se concentrer principalement sur des indicateurs quantitatifs comme les facteurs d'impact des publications. Comme le souligne le texte :

Ce système d'évaluation, qui s'appuie souvent sur la quantité de publications dans des revues à facteur d'impact élevé et sur la quantité de citations, [...] tend à sous-estimer la valeur des autres contributions, à réduire la reproductibilité des travaux de recherche et à freiner l'engagement des chercheurs en matière de partage, d'ouverture et de collaboration.

(Comité pour la science ouverte, 2022)

L'Appel de Paris encourage ainsi les institutions de recherche, les organismes de financement et les éditeurs académiques à adopter des approches plus holistiques et transparentes dans l'évaluation de la recherche. Ces approches devraient prendre en compte une diversité de produits de recherche, y compris les données, les logiciels, et autres contributions qui ne se limitent pas aux articles de revues. En outre, l'appel souligne l'importance de l'inclusion des contributions à l'enseignement, à l'engagement public, à l'impact sur les politiques, et à la collaboration internationale. Il invite à la création d'un écosystème de recherche plus équilibré et inclusif, qui reconnaît et récompense un éventail plus large de contributions scientifiques, facilitant ainsi une transition vers des pratiques de recherche plus collaboratives et ouvertes.

Les idées formulées dans l'Appel de Paris sont reprises au niveau européen la même année. Le Conseil de l'Union européenne adopte les conclusions du Conseil sur l'évaluation de la recherche et la mise en œuvre de la science ouverte. Ces conclusions encouragent les États membres à développer des systèmes d'évaluation de la recherche qui valorisent la transparence, l'accès ouvert, et la contribution sociétale de la recherche, alignant ainsi les politiques de recherche européennes avec des objectifs de science ouverte. En adoptant ces conclusions, l'Union européenne confirme son engagement à soutenir une transformation des pratiques de recherche dans la continuité des principes développés par les précédentes initiatives.

Toujours en Europe, les conclusions du Conseil sur la Publication universitaire de qualité, transparente, ouverte, fiable et équitable adoptées par le Conseil de l'Union européenne en 2023, affirment la nécessité d'un accès libre et immédiat pour les publications de recherche financées totalement ou en partie par des fonds publics.

Le récit de ces initiatives s'arrête ici en 2023, mais le mouvement mondial vers l'accès libre continue de se développer à travers de nombreuses autres actions. Ce nouveau paradigme transforme en profondeur la communication scientifique en rendant les résultats de la recherche accessibles à tous, sans barrières financières ou institutionnelles. Il repose sur des principes d'ouverture, de partage et de transparence, favorisant une diffusion plus équitable des connaissances. Nous espérons que la lecture de cet historique suscite chez le lecteur l'intérêt de se

documenter sur les nouvelles initiatives qui continueront de façonner l'avenir de l'accès libre à la connaissance scientifique. Cependant, ce changement majeur ne va pas sans susciter des réactions contrastées...

1.2.8. *Les voix contraires*

Un tel changement de paradigme dans la communication scientifique remet inévitablement en question la position dominante des éditeurs scientifiques traditionnels. Ces derniers, qui ont longtemps contrôlé l'accès aux recherches et bénéficié de modèles économiques lucratifs basés sur les abonnements et les frais de publication, se retrouvent confrontés à un modèle qui favorise l'accès libre et immédiat aux résultats de recherche. Bien que les grandes maisons d'édition n'aient pas toujours pris position officiellement contre la science ouverte, leur réponse peut être perçue comme nuancée ou même réticente. Elles développent des modèles hybrides ou des accords transformants, qui peuvent être vus comme des tentatives de s'adapter tout en préservant leur rôle central et leurs bénéfices. Cette réticence implicite à embrasser pleinement les principes de la science ouverte soulève des inquiétudes quant à la persistance de barrières financières et d'inégalités d'accès à l'information scientifique, malgré les progrès vers plus de transparence et d'accessibilité.

De l'autre côté du spectre, avec des métriques et des évaluations plus transparentes et accessibles, ainsi que des attentes accrues pour le partage des données de recherche, certains chercheurs craignent une augmentation de la surveillance sur leur travail. Cette crainte du « flicage », héritée du paradigme du *Publish or Perish* (Bouchard & Boudry, 2024), qui a longtemps prévalu, met en lumière les défis éthiques et psychologiques associés à une plus grande ouverture, nécessitant une gestion délicate pour encourager l'adoption sans renforcer la pression déjà lourde sur les scientifiques. Concernant ce point, selon nous, l'utilisation de l'accès libre dans le processus d'évaluation est inéluctable et se structure d'ores-et-déjà. Cependant, cela représente une opportunité singulière pour le chercheur. À l'inverse du dictat implicite des éditeurs avec les cadres d'évaluation imposés par le facteur d'impact, il est désormais l'acteur principal de ces nouveaux processus d'évaluation qui se veulent plus justes et inclusifs. Nous abordons cette problématique et apportons une réponse au cours du prochain chapitre.

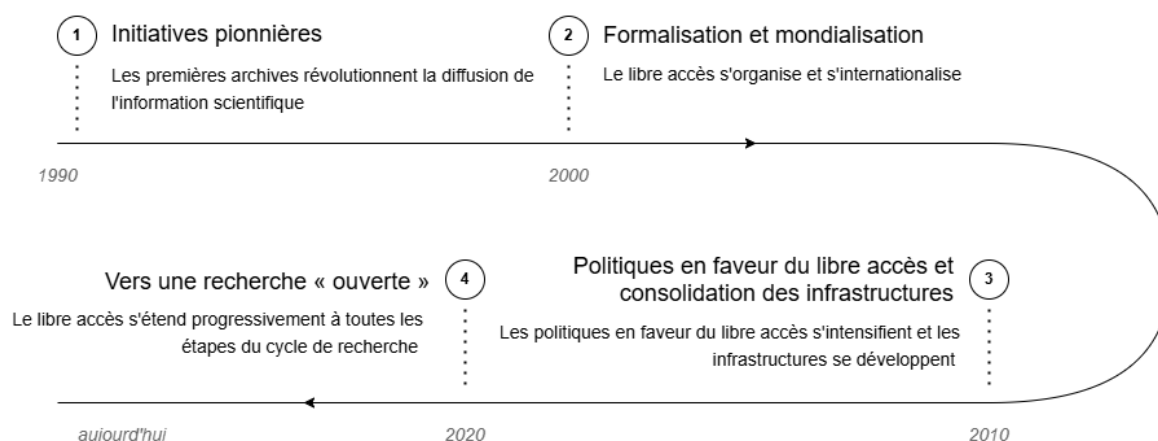
1.2.9. *Conséquences d'une appropriation massive*

L'historique développé ci-dessus, bien que dense, est nécessaire pour appréhender la diversité et la richesse des initiatives qui composent le mouvement pour le libre accès. Cette multitude d'initiatives illustre cette progression constante vers un accès libre généralisé, soutenu par des

politiques nationales et internationales, des infrastructures robustes et un engagement croissant de la communauté scientifique mondiale. Pour une compréhension plus concise de ce chemin, nous regroupons ces initiatives en plusieurs grandes catégories présentées en figure 1 et explicitées ci-après.

Figure 1

Les grandes étapes du développement du libre accès



Initiatives pionnières

Au début des années 1990, le mouvement du libre accès prend racine dans un contexte de digitalisation croissante de la communication scientifique. Cette période est marquée par l'émergence de projets pionniers qui visent à exploiter les nouvelles capacités offertes par l'Internet pour la diffusion rapide et ouverte des connaissances scientifiques. La création d'arXiv en 1991 par Paul Ginsparg, initialement destinée aux physiciens, est un jalon important. Ce projet exemplifie la capacité du numérique à catalyser la diffusion des découvertes, ouvrant la voie à des pratiques similaires dans d'autres disciplines.

L'impact d'arXiv encourage rapidement l'adoption de modèles analogues dans d'autres domaines spécialisés. Par exemple, PubMed Central, lancé à la fin des années 1990, se positionne comme un pivot pour les sciences de la vie, proposant un accès libre et sans entrave aux publications de recherche biomédicale. Cette période est également témoin de la diversification des initiatives selon les disciplines, chaque communauté adaptant les outils de libre accès à ses besoins spécifiques.

Ces premières initiatives démontrent une réceptivité variée au sein de la communauté scientifique, où certaines disciplines, telles que la physique et la biologie, adoptent rapidement ces nouveaux

outils, tandis que d'autres, comme les sciences humaines, restent plus réticentes en raison de préoccupations liées aux droits d'auteur et à la qualité de la relecture par les pairs.

Le paysage du libre accès de cette décennie est caractérisé par une exploration des possibilités offertes par le numérique, posant les fondations pour des développements futurs. Cette phase d'initiation révèle non seulement l'ingéniosité des chercheurs à s'approprier les outils numériques mais souligne aussi les défis inhérents à une adoption uniforme du libre accès à travers les différentes disciplines scientifiques.

Formalisation et mondialisation du libre accès

À l'aube du nouveau millénaire, le libre accès connaît une phase de formalisation de ses principes et de son expansion à l'échelle internationale. Cette période est marquée par des efforts concertés pour établir des standards clairs et des directives pour le libre accès, catalysant ainsi son adoption globale. En 2002, la Budapest Open Access Initiative établit un cadre formel pour le libre accès, définissant clairement les objectifs et les méthodes pour rendre la littérature scientifique librement accessible en ligne. Cette initiative est rapidement suivie par les déclarations de Bethesda et de Berlin en 2003, qui renforcent les principes du libre accès et encouragent les institutions et les gouvernements à adopter des politiques en sa faveur.

Durant cette décennie, le libre accès s'étend au-delà de ses foyers initiaux. Des initiatives comme SciELO en Amérique Latine montrent comment le libre accès peut s'adapter aux contextes culturels et régionaux spécifiques, augmentant la visibilité internationale de la recherche produite dans ces régions. Ces développements mettent en lumière les variations culturelles et étatiques dans l'adoption du libre accès, illustrant une tendance à une appropriation diversifiée du mouvement selon les différentes aires géographiques et les cadres politiques.

Cette époque est également témoin de la croissance de l'infrastructure de soutien au libre accès, avec la création de dépôts institutionnels et la mise en place de politiques par des entités telles que le National Institutes of Health aux États-Unis, qui exige désormais un dépôt libre accès des recherches qu'il finance. En France, l'archive ouverte Hyper Articles en Ligne (HAL) devient une référence majeure pour diverses disciplines, offrant une plateforme robuste pour l'auto-archivage des publications scientifiques. @rchiveSIC est lancée peu après pour répondre spécifiquement aux besoins des sciences de l'information et de la communication, soulignant ainsi l'adoption disciplinaire variée du libre accès.

Ainsi, cette période de formalisation et d'expansion globale solidifie les fondations du libre accès, tout en révélant et en adressant les défis de son adoption universelle.

Politiques en faveur du libre accès et consolidation des infrastructures

Le libre accès entre dans une phase de consolidation à partir de 2010, avec une focalisation renforcée sur le renforcement des infrastructures et l'intégration des politiques de soutien. La Commission Européenne, avec son plan Horizon 2020, joue un rôle clé en exigeant la publication en libre accès de toutes les recherches qu'elle finance. Cette politique est un moteur important pour l'adoption du libre accès en Europe et encourage le développement de plateformes comme OpenAIRE, qui facilite le partage et la visibilité des recherches financées par des fonds publics européens.

Au cours de cette période, les bibliothèques universitaires s'impliquent davantage dans la promotion du libre accès, en développant des services de soutien à l'auto-archivage et en négociant des contrats avec des éditeurs pour sécuriser l'accès aux publications. Les infrastructures comme les dépôts institutionnels se multiplient, soutenant ainsi une diversification des actions visant à rendre la recherche librement accessible.

Dans le paysage du libre accès, les politiques du Royaume-Uni et de la France illustrent deux approches distinctes. Le Royaume-Uni, influencé par le rapport Finch, a adopté une stratégie favorisant la voie dorée, soutenant financièrement les frais de traitement d'article pour encourager la publication dans des revues en libre accès. Cette approche a été institutionnalisée par des organismes de financement comme le Research Councils UK, qui remboursent les APC pour les travaux qu'ils financent. En contraste, la France privilégie la voie verte et préconise le dépôt libre des publications dans des archives ouvertes comme HAL, sans frais pour les auteurs (Arènes et al., 2022). Cette différence reflète des choix stratégiques divergents dans le financement et la gestion des publications scientifiques, marquant une préférence nationale pour des modèles économiques spécifiques du libre accès.

La consolidation des infrastructures et des politiques durant cette période démontre un engagement croissant de la communauté scientifique envers le libre accès, signe incontestable de la maturation du mouvement.

Vers une recherche « ouverte »

Alors que le libre accès s'impose de plus en plus comme un pilier fondamental dans les politiques de recherche à travers le monde, les principes que le mouvement porte se diffusent à toutes les

étapes du processus de création scientifique, appelant à de nouveaux changements de paradigme dès les années 2016-2018.

Cette période met également en lumière des défis spécifiques, tels que les coûts importants associés aux modèles de publication basés sur les frais de traitement d'articles. La réaction à ces obstacles suscite une réflexion accrue sur le modèle économique des publications en libre accès, en particulier concernant les frais de traitement d'articles. Parallèlement, la nécessité de standards interopérables devient plus pressante pour assurer une meilleure intégration des différentes plateformes et archives de recherche, un enjeu crucial pour l'accès ouvert.

La pandémie de COVID-19 accélère cette dynamique d'intégration du libre accès en démontrant la nécessité vitale d'un accès rapide et sans entrave à l'information scientifique pour répondre efficacement aux crises mondiales. Elle stimule également l'adoption de pratiques de libre accès par des éditeurs, tels que Elsevier, Springer Nature et Wiley, et des institutions qui ouvrent temporairement leurs ressources pour soutenir la recherche et l'innovation en temps de crise.

Cette période démontre la complexité de la transition vers une science totalement ouverte et accessible et reflète une nécessaire évolution du libre accès dans les années à venir. À l'image de Rome qui ne s'est pas faite en un jour, la transformation vers un modèle de recherche entièrement « ouvert » s'avère être un processus long et complexe qui ne cesse d'évoluer.

Ces quelques lignes illustrent le fait que la notion de libre accès, et par extension de science ouverte, ne se limite pas au simple dépôt de la publication scientifique sur une plateforme dédiée. Ce changement de paradigme de diffusion bouleverse et s'impose dans l'ensemble des activités constitutives du processus de création scientifique présentées en figure 2, de la préparation du projet de recherche à l'évaluation même de l'impact du chercheur. Nous détaillons ci-après comment ce changement accompagne la science vers un idéal de collaboration globale.

Dans la phase de préparation, le libre accès révolutionne la manière dont les priorités de recherche sont définies et partagées. Grâce à l'ouverture des données, des publications, et même des prépublications, les chercheurs ont un accès direct à une vaste quantité d'informations en amont de leurs projets. Cette disponibilité immédiate permet une revue de la littérature plus exhaustive et dynamique, offrant ainsi une meilleure identification des lacunes de la recherche actuelle. Les plateformes de libre accès permettent aussi de consulter les plans de recherche d'autres projets, favorisant ainsi la convergence des efforts scientifiques autour des

problématiques les plus urgentes. Cette transparence facilite non seulement la définition des priorités, mais encourage également une coordination renforcée entre les disciplines et les pays, ouvrant la voie à des collaborations interdisciplinaires et internationales. Ces collaborations, autrefois limitées par les barrières géographiques ou institutionnelles, deviennent plus fluides et permettent la création de projets de recherche plus ambitieux, diversifiés et inclusifs, capables de répondre aux défis mondiaux complexes.

La phase de découverte se trouve également transformée par cette logique d'ouverture. L'accès libre à l'information scientifique, qu'il s'agisse d'articles, de données brutes, de logiciels, ou de codes sources, non seulement élargit le réservoir de connaissances disponibles, mais en accélère aussi la circulation. Cela permet aux chercheurs de bâtir plus rapidement de nouvelles connaissances sur des bases solides et validées par la communauté scientifique. L'essor des bibliothèques numériques, des dépôts de données ouverts, et des archives institutionnelles offre une accessibilité sans précédent aux travaux scientifiques, y compris ceux des chercheurs issus de régions sous-représentées. De plus, l'émergence de nouveaux outils numériques et les systèmes de gestion de l'information scientifique permettent de traiter et d'exploiter cette immense volumétrie d'informations de manière plus efficace. Ces outils améliorent considérablement le processus de veille scientifique, en facilitant la détection des tendances émergentes et des avancées les plus récentes dans divers domaines de recherche, tout en favorisant l'innovation et la création de nouvelles hypothèses de travail.

L'analyse des données est également profondément impactée par le libre accès. Les chercheurs ont désormais la possibilité d'accéder à des jeux de données massifs et diversifiés, publiés par d'autres chercheurs ou institutions, qu'ils peuvent réutiliser pour leurs propres travaux. Cette réutilisation n'est pas seulement un gain de temps, mais elle permet aussi d'augmenter la robustesse des analyses en croisant des données provenant de multiples sources. Par ailleurs, la culture de partage promue par la science ouverte encourage les chercheurs à rendre publics leurs propres protocoles, méthodes d'analyse, et résultats intermédiaires. Cette transparence accrue contribue à une plus grande rigueur scientifique, car elle permet une validation externe des résultats et une réplication des études par d'autres équipes de recherche. De plus, l'émergence de plateformes collaboratives pour le partage de données brutes et de protocoles, comme les data journals et les dépôts en ligne, renforce l'idée de la recherche comme un processus collectif, où chaque étape, de la collecte à l'analyse des données, peut être partagée et revue par la communauté.

Dans la continuité du partage de ces données, le processus d'écriture se trouve également transformé par les principes de la science ouverte. Les chercheurs sont encouragés à adopter une approche plus collaborative, où l'écriture scientifique devient un exercice collectif. Grâce aux plateformes de rédaction ouvertes, les manuscrits peuvent être coécrits, commentés et révisés en temps réel par plusieurs chercheurs, facilitant ainsi les échanges d'idées et l'amélioration continue du contenu. Cette collaboration élargie, souvent internationale, permet de réduire les biais et d'enrichir la qualité des publications. De plus, les brouillons, les prépublications et les versions intermédiaires des travaux peuvent être partagés avec la communauté scientifique avant même la soumission finale, offrant ainsi des opportunités de retour critique et de validation précoce par les pairs. Ce processus itératif renforce non seulement la rigueur scientifique, mais aussi l'accessibilité et la transparence des connaissances en construction.

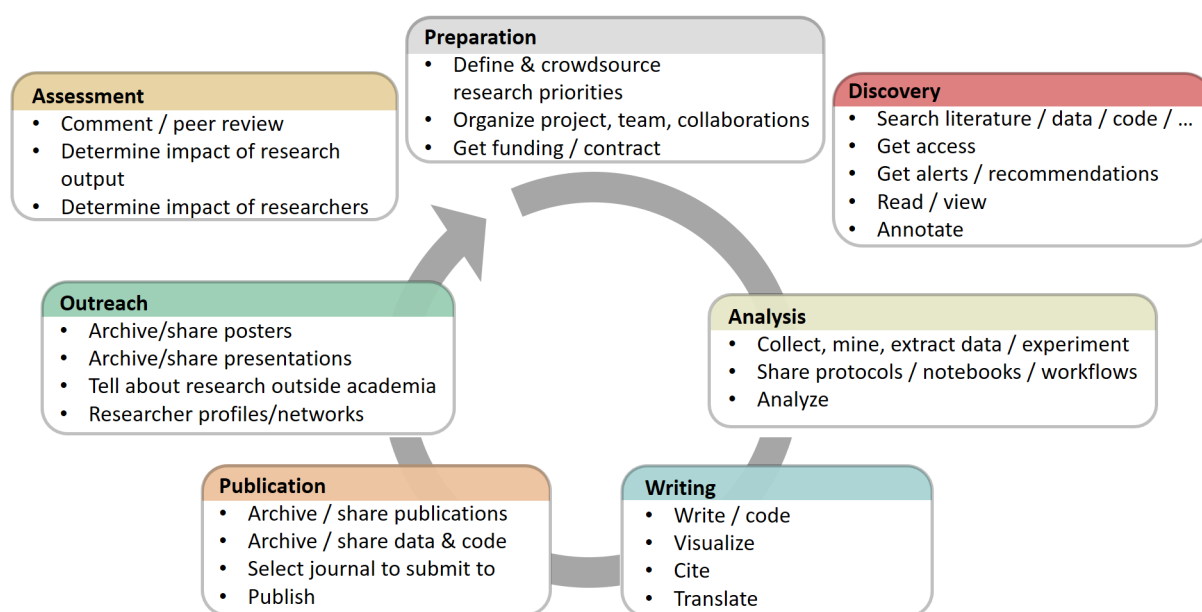
Les processus de publication et de diffusion, traditionnellement limités par les abonnements payants aux revues, sont fondamentalement redéfinis par le libre accès. L'édition scientifique, auparavant dominée par un petit nombre de grandes maisons d'édition, opérait dans un modèle oligopolistique, où l'accès aux connaissances était largement dicté par des considérations commerciales. Désormais les chercheurs peuvent publier leurs travaux sous leurs diverses formes sur des plateformes ouvertes et dédiées. Ces plateformes, à savoir les archives ouvertes et les revues en libre accès, permettent une disponibilité immédiate des résultats pour l'ensemble de la communauté scientifique, ainsi que pour le grand public. Cette accessibilité favorise une diffusion plus large et plus rapide des connaissances, augmentant l'impact et la visibilité des recherches au-delà des cercles académiques traditionnels. De plus, le libre accès permet de surmonter les barrières financières qui pouvaient limiter l'accès à l'information scientifique, en particulier pour les chercheurs des pays en développement ou pour ceux sans affiliation institutionnelle. Ce changement de paradigme contribue à démocratiser la science, en permettant à un plus grand nombre de personnes d'accéder aux dernières avancées scientifiques, de les exploiter et de participer à la création de nouvelles connaissances.

Enfin, la phase d'évaluation du travail scientifique est elle aussi profondément impactée par la science ouverte. L'ouverture des publications et des données permet une évaluation plus qualitative que quantitative, intégrant ainsi une diversité de perspectives et de critères allant bien au-delà du traditionnel facteur d'impact. Avec l'adoption de nouvelles métriques alternatives (altmetrics), l'impact des chercheurs est désormais mesuré non seulement par le nombre de publications, mais aussi par la manière dont leurs travaux sont diffusés, discutés, et réutilisés dans la communauté scientifique et au-delà. Comme l'expliquent Priem et al. (2010), « les altmetrics

offrent une perspective plus complète et nuancée de la manière dont les travaux scientifiques sont reçus et utilisés par différentes communautés ». Par exemple, les citations des données, les mentions dans les réseaux sociaux, les téléchargements, et les réutilisations dans d'autres contextes sont autant de nouveaux indicateurs de l'influence réelle d'un chercheur. Par ailleurs, l'évaluation par les pairs, processus central de la validation scientifique, bénéficie également de la transparence accrue offerte par le libre accès. Les revues pratiquant l'open peer review permettent non seulement de rendre les rapports d'évaluation accessibles, mais aussi d'attribuer un crédit aux évaluateurs, renforçant ainsi la reconnaissance du travail d'examen critique. Ce modèle favorise une science plus ouverte, plus rigoureuse, et plus inclusive, alignée avec les valeurs d'ouverture globale de la recherche scientifique.

Figure 2

Cycle du processus de recherche (Bezjak et al., 2018)



La taxinomie de la science ouverte définie par Pontika et Knoth (2015) et présentée en figure 3 appose des mots sur les notions développées précédemment et qui interviennent dans le cycle de la recherche. Ils répertorient l'intégralité des concepts associés à ce mouvement. Nous constatons, à nouveau, que derrière ces deux mots se cache une trentaine de concepts qui possèdent chacun des caractéristiques uniques, rendant leur maîtrise aussi enrichissante que complexe. Ce travail de clarification s'inscrit également dans une dynamique institutionnelle plus large, la plateforme Loterre, développée par l'Institut de l'information scientifique et technique

(INIST) du CNRS, propose un thésaurus structuré de la science ouverte qui détaille et hiérarchise les différents concepts mobilisés, en précisant leurs relations sémantiques (Institut de l'information scientifique et technique (Inist) - CNRS/UAR76, 2021).

Cette structuration taxinomique, bien qu'indispensable à la lisibilité du champ, n'est pas sans soulever des enjeux épistémologiques. En effet, classifier, hiérarchiser, établir des relations sémantiques, ce n'est pas seulement rendre compte du réel, c'est aussi le configurer. Michel Foucault, dans *Les Mots et les Choses* (1966), le rappelle avec force dans un passage devenu classique :

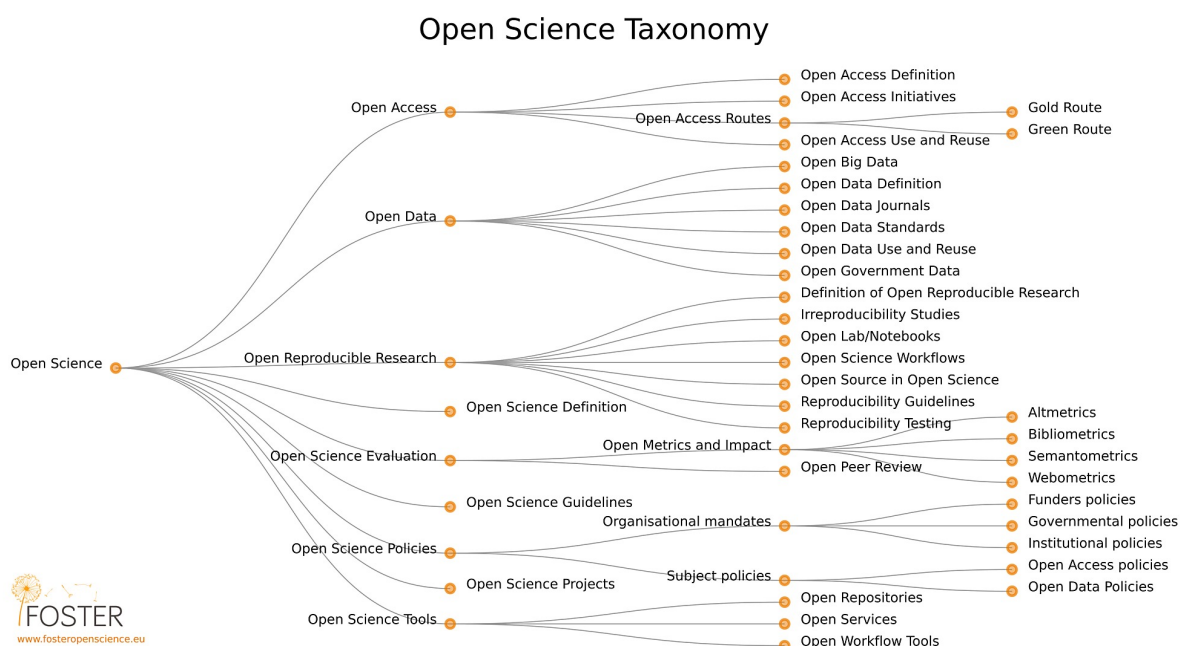
L'ordre des choses n'est pas un ordre arbitraire, et il n'est pas non plus donné une fois pour toutes dans la nature des êtres : c'est un ordre historique, qui régit les pratiques discursives, et fonde les modes de connaissance ; il impose à ce qui est vu et à ce qui est dit les conditions de leur visibilité et de leur énonciabilité. Il n'y a pas de savoir sans une certaine mise en ordre du visible et du dicible.

(Foucault, 1966)

À la lumière de cette perspective, les outils comme Loterre ou SoVisu ne sont pas de simples instruments techniques : ils participent à la production d'un certain ordre du savoir, en délimitant ce qui peut être nommé, relié, visualisé, donc reconnu comme scientifique et légitime dans le cadre de la science ouverte.

Figure 3

La taxinomie de la science ouverte (Pontika & Knoth, 2015)



Dans cette thèse, nous nous attachons à traiter principalement de la question des archives ouvertes. Derrière ce concept, nous approchons également et inévitablement les notions connexes de politiques gouvernementales et institutionnelles et de voies vers le libre accès. Pour rappel, nous distinguons deux types majeurs de libre accès. Premièrement la voie **dorée** implique la publication des travaux directement en libre accès sur le site de l'éditeur selon diverses modalités. Cette publication peut impliquer le paiement de frais de traitement d'article par l'auteur, une pratique de plus en plus critiquée en raison de son coût élevé. Par exemple, chez Nature, les options de publication peuvent entraîner des frais s'élevant à 10 290 euros, un montant conséquent (*Publishing Options* | Nature, s. d.). Cette méthode est aussi qualifiée de voie hybride. Cette voie peut aussi inclure des publications dans des revues subventionnées par des organismes, qui couvrent ces frais. Parallèlement, la voie **verte** repose, elle, sur l'auto-archivage des travaux de recherche par l'auteur, dans une archive ouverte, souvent après le respect d'une période d'embargo. Piwowar et al. (2018) proposent une classification plus développée des différentes voies vers le libre accès, bien que certaines soient moins répandues, et déclinent également celles décrites ici. Ils caractérisent, notamment, la voie **bronze** par des articles disponibles en accès ouvert mais sans licence formelle définissant les conditions de réutilisation. La voie verte nous intéresse particulièrement dans ce travail. Ce paradigme de diffusion, qui se

rapproche, d'une certaine façon, des échanges épistolaires originels, semble privilégié en France au travers des différents plans nationaux pour la science ouverte. Il repose directement sur l'utilisation des archives ouvertes, à savoir, en France, la plateforme HAL, qui s'avère être un terrain d'étude très riche et particulièrement intéressant à analyser.

1.3. Les archives ouvertes

Les archives ouvertes constituent un pilier central du libre accès à l'information scientifique. Elles offrent aux chercheurs un espace de dépôt où partager librement leurs travaux, favorisant ainsi une diffusion élargie et une accessibilité accrue du savoir. Contrairement aux modèles traditionnels de publication, ces dispositifs reposent sur une logique non commerciale et collaborative où les auteurs conservent le contrôle de leurs productions, tout en les rendant accessibles à la communauté. En France, l'archive ouverte HAL occupe une place structurante dans cet écosystème. Décrite par Charnay et Michau (2007) comme une « plate-forme commune inter-établissements d'archives ouvertes », HAL incarne la volonté des institutions publiques de se réapproprier les moyens de diffusion scientifique face à la domination croissante des éditeurs commerciaux (Chartron, 2010). Au-delà de la simple démocratisation de l'accès à l'information, les archives ouvertes modifient en profondeur les pratiques académiques. Comme le souligne Guédon (2008), elles instaurent de nouvelles normes de collaboration et d'évaluation scientifique. L'auto-archivage accélère la circulation des résultats et confère aux chercheurs un pouvoir accru sur la visibilité de leurs travaux. C'est dans cette perspective qu'il convient désormais de resserrer le regard. En passant de l'analyse générale des principes du libre accès à l'examen d'une infrastructure nationale spécifique, à savoir HAL, il s'agit de saisir comment ces enjeux globaux se traduisent concrètement dans un dispositif particulier, aux usages et à la gouvernance situés. Cette section propose d'explorer ces dynamiques à travers une analyse approfondie de la plateforme HAL, du rôle des métadonnées, et des spécificités propres aux sciences de l'information et de la communication.

1.3.1. L'archive ouverte HAL

L'archive ouverte HAL s'est progressivement imposée comme un dispositif central de la science ouverte en France, évoluant au gré des transformations stratégiques impulsées par les institutions de recherche, mais aussi en réponse aux besoins et aux retours de la communauté scientifique.

Le travail de Charnay et Michau (2007) dresse un descriptif technique d'époque de l'archive ouverte HAL. Ils décrivent le principe essentiel du dispositif, qui consiste à permettre à

l'utilisateur de télécharger le texte intégral d'un document scientifique assorti de certaines métadonnées, pour certaines facultatives, pour d'autres obligatoires. Il paraît nécessaire de décrire brièvement le processus de dépôt sur HAL ainsi que la nature des métadonnées collectées, dans la mesure où ces éléments constituent le socle des dynamiques d'archivage que nous analysons par la suite. Le dépôt consiste, pour l'auteur identifié, à renseigner un ensemble de métadonnées structurées, et, dans la plupart des cas, à téléverser le document scientifique en texte intégral. Toutefois, ce dernier n'est pas obligatoire : certains dépôts ne comportent qu'une notice bibliographique, ce qui révèle des usages différenciés du dispositif. Les métadonnées collectées incluent notamment des informations bibliographiques, à savoir, le titre, les auteurs, les différentes dates afférentes au document et le type, des affiliations institutionnelles normalisées, des éléments de classification disciplinaire, ainsi que, le cas échéant, des métadonnées spécifiques à une collection ou à un portail. Certaines d'entre elles sont obligatoires pour valider le dépôt, tandis que d'autres restent facultatives, laissant une marge d'interprétation et d'engagement à l'auteur du dépôt. Ces données s'appuient majoritairement sur des référentiels contrôlés (structures, revues, disciplines), afin de garantir la qualité, la cohérence et l'interopérabilité des dépôts dans l'écosystème de la science ouverte. Ce processus, bien que relativement stable dans ses grandes lignes, a connu certaines évolutions au fil du temps, à mesure que la plateforme HAL s'est développée et que les exigences en matière de qualité des données et d'interopérabilité se sont accrues. Comme le soulignent également les auteurs, le renseignement de ces métadonnées est essentiel pour permettre l'indexation automatique des contenus dans les moteurs de recherche scientifique et garantir la visibilité des travaux. C'est précisément dans ce dispositif que s'ancrent les enjeux de description, de visibilité et d'effort contributif qui structurent la problématique de cette thèse.

À l'occasion des 20 ans de HAL, le Centre pour la Communication Scientifique Directe (2021) publie une brochure qui retrace avec précision les étapes clés de l'évolution de la plateforme. La frise chronologique, présentée en figures 4 et 5, en est extraite. Cette frise ne se limite pas à une suite de jalons techniques. Elle révèle une trajectoire structurante : de la construction initiale d'une infrastructure d'auto-archivage à la constitution d'un véritable écosystème de médiation scientifique. Elle met en lumière la montée en puissance des dispositifs d'accompagnement (nouveaux portails, ergonomie de dépôt, ambassadeurs), l'institutionnalisation de HAL en tant qu'infrastructure stratégique, ainsi que le virage vers la découvrabilité mesurable avec, par exemple, HAL Monitor (CCSD, 2025) ou encore HALiance (CCSD, 2022). Ces évolutions confirment que l'auto-archivage est aujourd'hui soutenu par des dispositifs sociotechniques à la

fois humains et algorithmiques. Sa lecture permet de replacer certains jalons déjà évoqués dans une perspective temporelle cohérente, en écho à l'historique précédemment développé.

Les premières années sont marquées par la volonté d'intégrer des portails disciplinaires déjà existants, à l'instar d'ArXiv pour la physique, RePEc pour l'économie ou encore PubMed Central pour les sciences biomédicales. À partir de 2005, l'ouverture de HAL-SHS élargit considérablement le périmètre disciplinaire de la plateforme, en rendant possible l'entrée des sciences humaines et sociales dans la dynamique d'auto-archivage. Cette extension soulève plusieurs questions : comment les spécificités disciplinaires des SHS ont-elles influé sur les formes d'appropriation du dispositif ? Peut-on identifier des temporalités ou des modalités de dépôt différenciées, notamment dans le champ des SIC ? Ces interrogations seront approfondies dans la suite du travail.

L'évolution de HAL se caractérise également par une diversification progressive des formats accueillis (textes, images, vidéos, sons, codes sources), ainsi que par un processus de mutualisation technique au service des établissements. Dès 2006, cette logique se concrétise dans le champ des sciences de l'information et de la communication : @rchiveSIC, initialement autonome, est intégré à HAL, ce qui a pu contribuer à infléchir les dynamiques d'archivage au sein de la discipline. Cette évolution prolonge une réflexion amorcée dès les années 2000 sur la spécificité des SIC face aux dispositifs d'archivage scientifique, comme le montre la démarche pionnière de conception d'@rchiveSIC analysée par Gallezot et al. (2003).

Parallèlement, plusieurs ajustements techniques visent à simplifier l'expérience de dépôt : amélioration du formulaire, interconnexion avec des identifiants chercheurs, développement de fonctionnalités de moissonnage. À ces évolutions s'ajoute la mise en place d'une interface de programmation (API), qui facilite l'interopérabilité de HAL avec d'autres systèmes d'information et permet l'extraction automatisée de données. C'est sur cette infrastructure que repose la collecte des données analysées dans le cadre de cette thèse, et dont les modalités sont explicitées plus en aval de ce travail. Ces évolutions, souvent pensées comme des leviers d'adhésion, coïncident avec une hausse significative du nombre de dépôts et de contributeurs. Mais dans quelle mesure ces transformations techniques ont-elles effectivement facilité l'appropriation du dispositif par les chercheurs ? S'agit-il d'un tournant, ou d'un simple effet d'accélération dans un mouvement amorcé en amont ?

Au cours du troisième chapitre, nous reviendrons plus en détail sur ces différentes évolutions, en interrogeant les effets qu'elles peuvent sur les pratiques de dépôt, les formes d'appropriation du dispositif et les dynamiques d'engagement des contributeurs.

Figure 4

« Les temps forts de la vie de HAL », première partie (Centre pour la Communication Scientifique Directe, 2021)

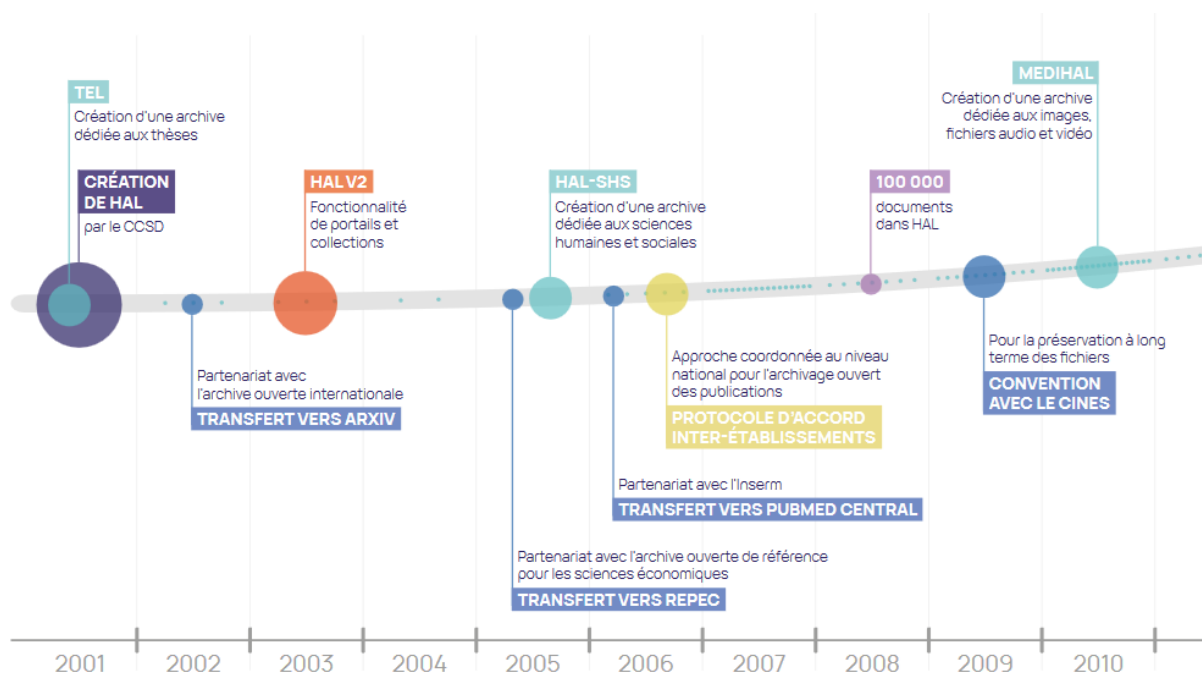
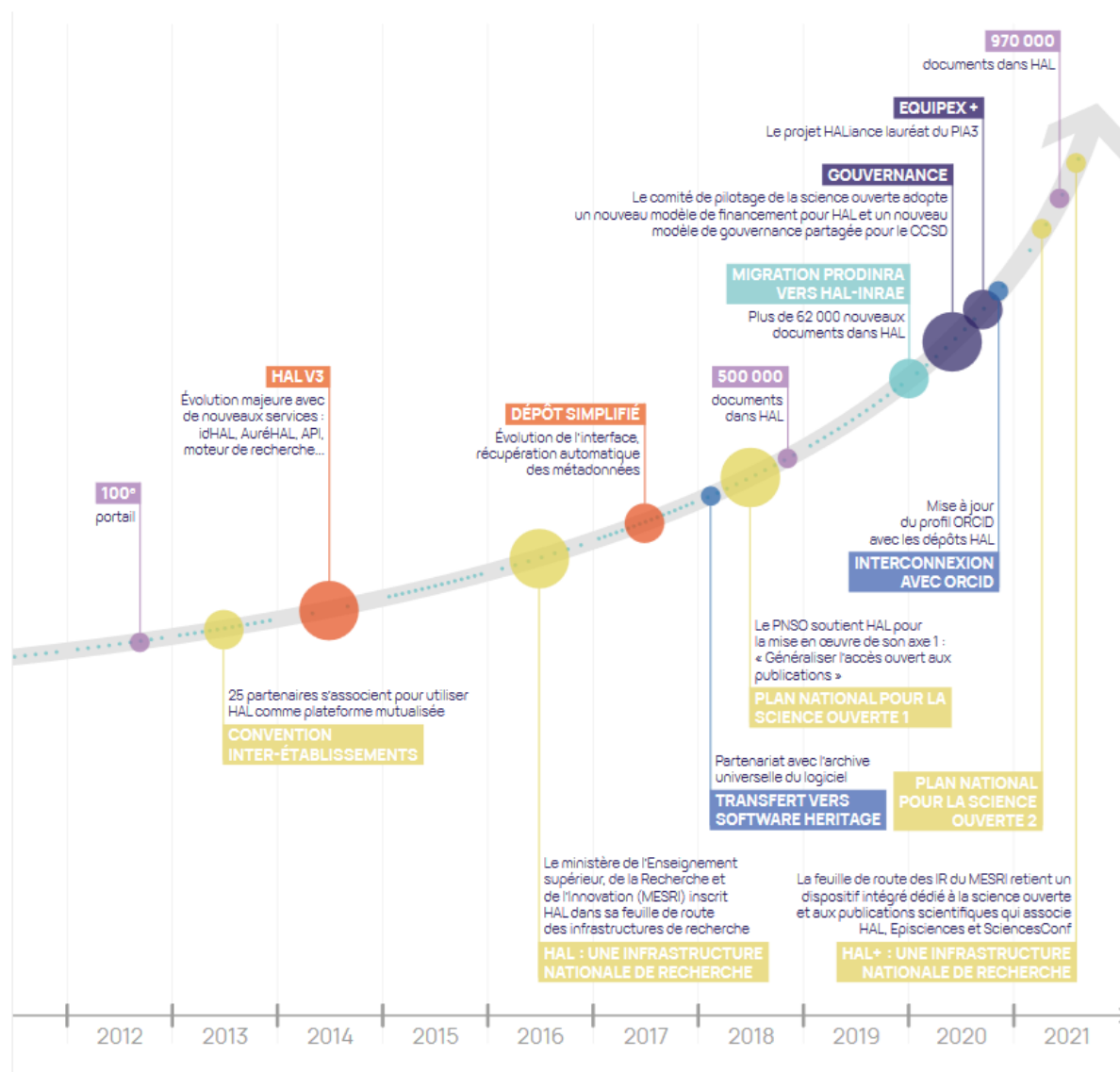


Figure 5

« Les temps forts de la vie de HAL », seconde partie (Centre pour la Communication Scientifique Directe, 2021)



Ces dynamiques d'appropriation ne doivent toutefois pas masquer une réalité plus diffuse, mais tout aussi structurante : les formes d'usages périphériques, de détournement ou de non-usage du dispositif. Car si HAL constitue une plateforme de référence, elle ne résume pas à elle seule l'ensemble des stratégies de diffusion et de mise en visibilité adoptées par les chercheurs comme le montre également l'étude de Schöpfel (2020), qui identifie une grande hétérogénéité dans l'usage de HAL par les unités de recherche de l'Université de Lille, selon les disciplines et la présence ou non de collections structurées. En parallèle ou en alternative au dépôt dans HAL, nombreux sont ceux qui mobilisent d'autres espaces de diffusion : plateformes

socio-académiques telles que ResearchGate ou Academia.edu, carnets de recherche sur Hypothèses, blogs scientifiques, voire sites personnels (Boudry & Bouchard, 2017). Ces canaux, bien que hors des cadres institutionnels de l'accès libre, sont parfois perçus comme plus souples, plus accessibles ou mieux adaptés à une communication scientifique individualisée. Ils permettent notamment de contourner certaines contraintes de dépôt, de mettre en scène un profil public, ou de diffuser plus librement des versions non officielles ou en cours de rédaction. S'ajoutent à cela des pratiques de dépôt sélectif : certains chercheurs choisissent de ne verser sur HAL que leurs productions les plus stratégiques ou les plus reconnues, tandis que d'autres omettent volontairement les documents issus de collaborations périphériques, ou encore les publications en cours d'évaluation. À l'inverse, le non-dépôt peut résulter d'un manque de temps, d'un défaut d'information, voire d'un désaccord de principe avec les logiques d'archivage centralisé. Ces logiques, souvent invisibles dans les métriques classiques, soulignent les limites de l'analyse strictement quantitative des contributions : ce que HAL ne montre pas, à savoir, les absents, les non-dits, les marges, participe pleinement à la compréhension des régimes de visibilité dans l'écosystème de la science ouverte. Leur prise en compte permet de réintroduire une dimension critique dans l'étude des dispositifs, en rendant visible ce qui échappe à l'archivage normé, mais qui participe néanmoins de la fabrique contemporaine de la communication scientifique.

Cette attention portée aux usages périphériques, aux tactiques de contournement et aux régimes de visibilité différenciés trouve une résonance particulière dans le champ des sciences de l'information et de la communication. Par leur ancrage interdisciplinaire et leur sensibilité aux conditions de production et de circulation des savoirs, les SIC se distinguent justement par une posture réflexive sur les dispositifs de médiation documentaire. C'est ce que nous allons maintenant explorer plus précisément.

1.3.2. L'archivage en sciences de l'information et de la communication

L'archivage revêt une dimension particulière dans les sciences de l'information et de la communication (SIC), tant en raison des caractéristiques épistémologiques de la discipline que de son rapport réflexif aux dispositifs de diffusion des savoirs. Cette posture réflexive rejoint les approches développées par Chaudiron & Ihadjadene (2010), qui plaident pour une analyse située des pratiques informationnelles, attentive aux contextes de production et aux médiations techniques et sociales impliquées dans les usages des dispositifs documentaires. Cette réflexivité se manifeste également dans les formes de résistance ou de contournement de HAL par les chercheurs en SIC, qui recourent parfois à d'autres modalités de publication comme les sites personnels, carnets de recherche, ou blogs, pour préserver la contextualisation de leurs travaux ou

la forme dialogique de leur écriture. Ces pratiques peuvent être analysées comme des formes de dissensus documentaires (Broudoux, 2017), révélant une tension entre injonction à la visibilité et préservation des logiques propres à la communication scientifique dans les SHS. Les SIC constituent un champ interdisciplinaire, aux frontières poreuses, où les objets d'étude, les méthodes et les formes de production scientifique sont particulièrement diversifiés. Cette hétérogénéité se manifeste dans les pratiques de dépôt : les chercheurs de la discipline peuvent aussi bien archiver des articles de revue que des actes de colloques, des communications, voire des supports non textuels issus de recherches-crédation ou d'enquêtes qualitatives. Ces formes diversifiées de diffusion rejoignent les résultats de l'enquête de Michel & Chekib (2015), qui montrent que les chercheurs en SHS arbitrent entre de multiples vecteurs de publication, combinant dispositifs formels et informels en fonction de leurs contraintes, de leurs cibles et de leur rapport au public; Par ailleurs, l'archivage en SIC ne se limite pas à une opération technique : il s'inscrit dans une posture critique, attentive aux conditions de production, de circulation et de réception des savoirs. À ce titre, les SIC sont doublement concernées par le développement de l'accès ouvert, à la fois comme communauté scientifique, actrice de la science ouverte, et comme discipline dont les objets portent sur les processus informationnels eux-mêmes (Chartron & Schöpfel, 2017). C'est selon cette perspective qu'est conçu le projet @rchiveSIC, archive ouverte francophone dédiée aux publications en SIC, lancée en 2002 à l'initiative de plusieurs laboratoires et soutenue par le CNRS. Conçue comme un outil de structuration disciplinaire, cette archive visait à remédier à la dispersion des publications, à offrir une plus grande visibilité aux travaux francophones, et à expérimenter de nouvelles formes de médiation scientifique, dans une logique de recherche-action (Gallezot et al., 2003). Dans le domaine des SIC en particulier, les archives ouvertes jouent ainsi un rôle crucial pour assurer la visibilité de travaux souvent peu représentés dans les bases de données commerciales. Prime-Claverie et Pouchot (Prime-Claverie & Pouchot, 2010) montrent, à travers une analyse comparative entre @rchiveSIC et la base Francis, que certaines thématiques propres aux SIC, telles que l'information scientifique et technique, l'économie des industries culturelles ou encore l'éducation aux médias, apparaissent de manière bien plus marquée dans les archives ouvertes que dans les systèmes commerciaux traditionnels. Leur étude met en lumière le potentiel des archives ouvertes non seulement comme outils de diffusion, mais aussi comme instruments de représentation fidèle et dynamique de la recherche en SIC. Mais cette posture critique et réflexive propre aux SIC influence-t-elle réellement les pratiques d'auto-archivage de ses chercheurs ? C'est l'une des questions que nous explorerons dans la troisième partie de ce travail, à travers une cartographie comparative des dépôts sur l'archive HAL.

1.3.3. Les métadonnées

Comme évoqué précédemment, les métadonnées jouent un rôle essentiel dans l'efficacité et l'utilité des archives ouvertes. Elles constituent la base de l'organisation, de la découverte et de l'interopérabilité des ressources scientifiques. Ce sont en effet elles qui sont moissonnées par les systèmes automatisés afin d'alimenter les dispositifs de recherche d'information à destination de la communauté scientifique. En ce sens, les métadonnées sont au cœur des principes FAIR (pour *Findable, Accessible, Interoperable, Reusable*) qui définissent les conditions d'une gestion optimisée des données scientifiques, c'est-à-dire organisée de manière à en favoriser la découverte, l'accès, la compatibilité technique et la réutilisation par d'autres acteurs de la recherche. Le premier pilier, *Findable*, repose précisément sur la capacité à rendre une ressource facilement repérable. Sans métadonnées structurées, normalisées, et surtout présentes, aucune trouvabilité n'est possible. Elles conditionnent donc l'accès, la réutilisation et la circulation des savoirs au sein des écosystèmes numériques de la recherche.

Au-delà de cette fonction de repérage essentielle à la construction et au partage des connaissances, les métadonnées soulèvent également des enjeux en matière de traçabilité et de visibilité de l'activité scientifique. Dès lors que l'archive HAL sert pour l'évaluation de la performance des universités, des organismes de recherche, des laboratoires et peut-être, des chercheurs eux-mêmes, comme en témoigne notamment le guide d'export de données produit par le Hcéres (2024), « l'exhaustivité, la pertinence et la qualité des données et métadonnées de HAL deviennent un enjeu majeur, pas seulement technique mais surtout politique et éthique » (Schöpfel et al., 2024). Dans cette perspective, le recours à des outils tels que ceux proposés par le Hcéres témoigne d'un basculement. Si l'on commence à se baser sur l'information scientifique archivée sur HAL pour évaluer la performance scientifique, selon une acception normative qui réduit souvent cette notion à la quantité de publications visibles et référencées., c'est bien que celle-ci regroupe désormais un volume significatif de travaux, constituant une base plutôt représentative de l'activité de recherche. À ce titre, nous choisissons de ne pas mobiliser ici les indicateurs du Baromètre de la Science Ouverte (Jeangirard, 2019), dont le périmètre méthodologique diffère sensiblement. Conçu pour mesurer la part de publications disponibles en accès ouvert à l'échelle nationale ou déclinés pour d'autres granularités plus précises, tous canaux confondus et ne permet pas de qualifier précisément les usages, la qualité descriptive ou les logiques de contribution spécifiques à une archive comme HAL, qui sont pourtant au cœur de notre analyse. Or, cette évolution, bien que encourageante, ne suffit pas à garantir le respect des principes FAIR. L'acte de déposer son travail sur le dispositif HAL peut se limiter au simple

renseignement de seulement quelques métadonnées. Le nombre croissant de dépôts sur l'archive ne dit alors rien de la qualité ni de la complétude des métadonnées qui en permettent la trouvabilité, l'interopérabilité et la réutilisation, alors qu'il s'agit là d'un des objectifs principaux poursuivis par le dispositif. Ce constat rejoint les préoccupations soulevées dans les travaux sur la documentarisation (Zacklad, 2004), où la qualité documentaire est envisagée non comme un attribut objectif, mais comme le résultat d'un processus sociotechnique structurant les conditions de circulation et de réception des productions sémiotiques. Cette approche rejoint également les travaux de Jean-Max Noyer (1995), qui rappelle que les données documentaires ne relèvent pas uniquement d'un traitement technique, mais participent d'un processus structurant d'évaluation et de représentation de l'activité scientifique, au croisement de la bibliométrie, de la scientométrie et de l'infométrie. Dans cette perspective, les métadonnées, entendues comme ressources terminologiques, ontologiques et attributs de description, ne sont pas de simples descripteurs neutres : elles traduisent une intentionnalité dans la production documentaire, et participent à l'orientation des usages. Leur absence ou leur incomplétude reflète ainsi des arbitrages, qu'ils soient explicites ou implicites, sur ce que le chercheur choisit de rendre visible, traçable, voire interprétable au sein des transactions communicationnelles. C'est précisément sur ce point que se concentrent les analyses développées au cours des deux chapitres suivants : au-delà des chiffres de l'ouverture, dans quelle mesure les pratiques d'archivage des chercheurs permettent-elles de rendre les productions réellement FAIR ? En d'autres termes, les métadonnées sont-elles suffisamment renseignées pour que l'idéal technique des FAIR rejoigne les réalités concrètes de la science ouverte ?

Avant d'apporter un éclairage à ces questions, il nous semble nécessaire de revenir sur une expérience concrète d'accompagnement à l'auto-archivage, initiée localement mais porteuse d'enjeux plus larges.

Certes, aborder une piste de solution avant même d'avoir établi un diagnostic peut sembler heuristiquement discutable. Pourtant, c'est bien à travers cette expérimentation que surgissent les interrogations évoquées précédemment. Ce renversement apparent de la logique démonstrative reflète en réalité le cheminement effectif de notre réflexion, tel qu'il s'est construit au fil des années. En exposant ici les origines, les choix et les évolutions du projet SoVisu, nous souhaitons rester fidèles à cette dynamique intellectuelle, où la pratique a précédé, nourri et orienté la problématisation théorique.

Partie 2 - SoVisu, un dispositif d'accompagnement à l'auto-archivage

« SoVisu comble une lacune de HAL en remplaçant le chercheur dans une posture active et réflexive sur l'ensemble de sa trajectoire documentaire. [...] Ce rôle élargi, qui va au-delà du simple dépôt, renforce la responsabilisation du chercheur quant à la qualité de son profil dans l'écosystème de la science ouverte. »

Dans ce second chapitre, nous retraçons la genèse d'un dispositif d'accompagnement à l'auto-archivage, développé au sein de l'Université de Toulon. En 2019, l'équipe présidentielle de l'université sollicite notre directeur de recherche, David Reymond, afin d'explorer la mise en place d'un système visant à améliorer la visibilité des compétences de recherche au sein de l'établissement (Reymond & Galliano, 2019). Le système doit poursuivre deux objectifs principaux : d'une part, offrir un outil de pilotage stratégique de la recherche scientifique à travers une cartographie interactive des compétences ; d'autre part, renforcer la synergie des collaborations internes et externes des chercheurs. C'est dans ce cadre que le projet prend forme. Alors stagiaire au sein du laboratoire de l'Institut Méditerranéen des Sciences de l'Information et de la Communication (IMSIC), j'ai pu, dès le début, contribuer activement à ce projet. Les premières lignes d'instructions, plutôt que les traditionnels mots, s'inscrivent dans une démarche de recherche-action et symbolisent ainsi mes premiers pas dans le domaine de la recherche universitaire et constituent un fil rouge qui guide les recherches réalisées au cours de cette thèse, présentées dans ce manuscrit. J'insiste sur l'expression « lignes d'instructions », car travailler sur un objet d'étude aussi étroitement lié à la technique constitue, à mon sens, une difficulté spécifique pour s'inscrire pleinement dans les cadres théoriques et méthodologiques propres aux sciences de l'information et de la communication et c'est pourtant le parti pris et la volonté que je me fixe ici. C'est pour cette raison que les chapitres de cette thèse sont agencés dans cet ordre, même si cela peut sembler déroutant au premier abord. Ce chapitre retrace donc des éléments de contexte propres au projet, les choix entrepris et l'évolution des fonctionnalités de la plateforme jusqu'en 2022, année où le projet prend une nouvelle envergure nationale grâce à la participation d'autres universités françaises de renom au sein du consortium CRISalid (Reymond, 2023 ; Reymond & Bretel, 2024 ; Reymond & Tabariès, 2024) .

2.1. Contexte

Ce projet, amorcé à l'initiative de l'équipe présidentielle de l'Université de Toulon à l'été 2019, s'inscrit dans un contexte plus large de promotion de la science ouverte et de valorisation des expertises universitaires (Reymond & Galliano, 2019). . Il poursuit deux objectifs principaux : d'une part, offrir un outil de pilotage stratégique de la recherche scientifique à travers une cartographie interactive des compétences et, d'autre part, renforcer la synergie des collaborations internes et externes des chercheurs.

Le projet de cartographie des compétences, tel que proposé par les auteurs du rapport, s'inscrit pleinement dans la science ouverte. Le projet s'appuie sur l'archive ouverte pluridisciplinaire

HAL, promue par le gouvernement comme pierre angulaire pour l'archivage et la diffusion de l'information scientifique en France. Cette archive joue un rôle clé dans la collecte des données, permettant non seulement d'en extraire des informations, mais également, selon nous, d'identifier des expertises. Chaque chercheur peut ainsi retrouver, via cet outil, une représentation de la production scientifique associée à son identifiant. Cette démarche responsabilise les chercheurs en les impliquant directement dans la qualité et l'exhaustivité de leur profil. Ils sont, via SoVisu, encouragés à mettre à jour et enrichir leurs contributions sur cette archive, afin d'assurer une représentation fidèle et complète des informations exploitées pour la cartographie des compétences.

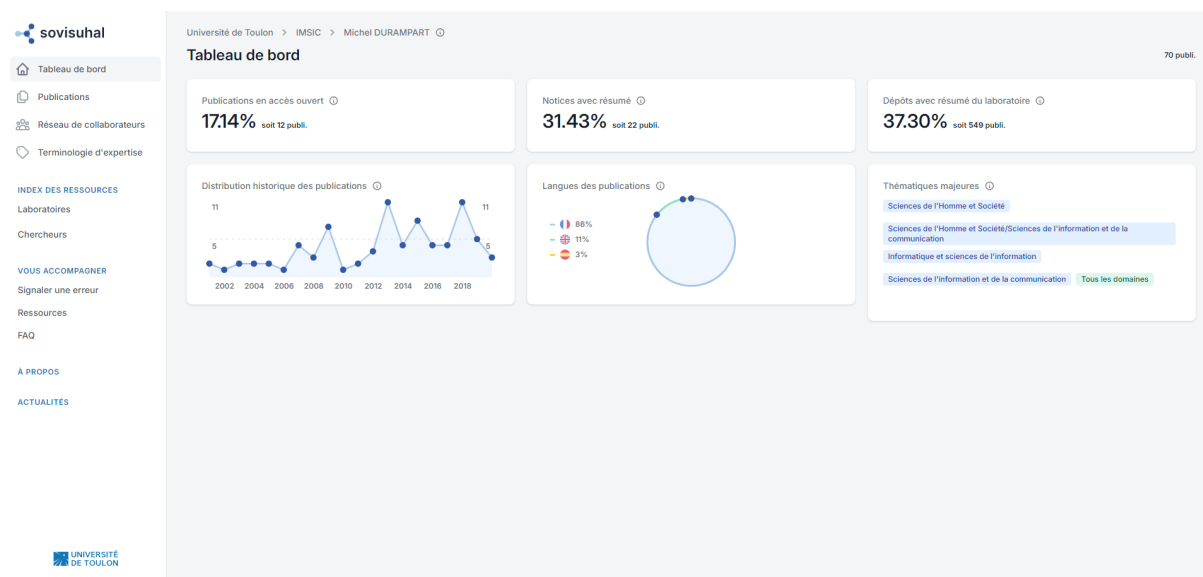
2.2. Un premier prototype fonctionnel

Le premier prototype, dénommé CartoHAL dans un premier temps, puis SoVisuHAL, élaboré conjointement avec mon directeur de recherche, dès les premiers mois de mon stage de recherche, se focalise sur les principales fonctionnalités demandées par la présidence. Nous ne détaillons pas ici les aspects techniques de cette première esquisse, car notre objectif est avant tout d'introduire le second prototype, plus abouti et pertinent. Ce premier prototype, développé en Python, s'appuie sur une liste d'identifiants HAL récoltés de manière semi-automatisée, à partir des noms et prénoms, depuis l'interface mise à disposition par l'archive. À partir de cette liste d'identifiants, nous procédons au moissonnage des notices associées en s'appuyant sur l'interface de programmation mise à disposition par HAL (*API HAL* | *API Archive Ouverte HAL*, s. d.). Ce processus permet de produire des visualisations synthétiques basées sur la bibliothèque D3.js (Bostock et al., 2011). Ces dernières offrent une représentation globale de la production scientifique d'un chercheur, mais également, par extension, celle du laboratoire auquel il appartient et, au-delà, de l'université. Le prototype propose quatre vues principales, chacune répondant à des objectifs spécifiques.

Une première vue (figure 6) synthétique regroupe des indicateurs clés calculés à partir des données récoltées. Elle permet à l'utilisateur d'obtenir une vision d'ensemble de l'état de la production scientifique archivée, incluant son évolution temporelle, la répartition linguistique des dépôts et une analyse des principales thématiques abordées. Ces dernières sont extraites des métadonnées associées aux dépôts, offrant ainsi une lecture à la fois quantitative et qualitative des travaux recensés.

Figure 6

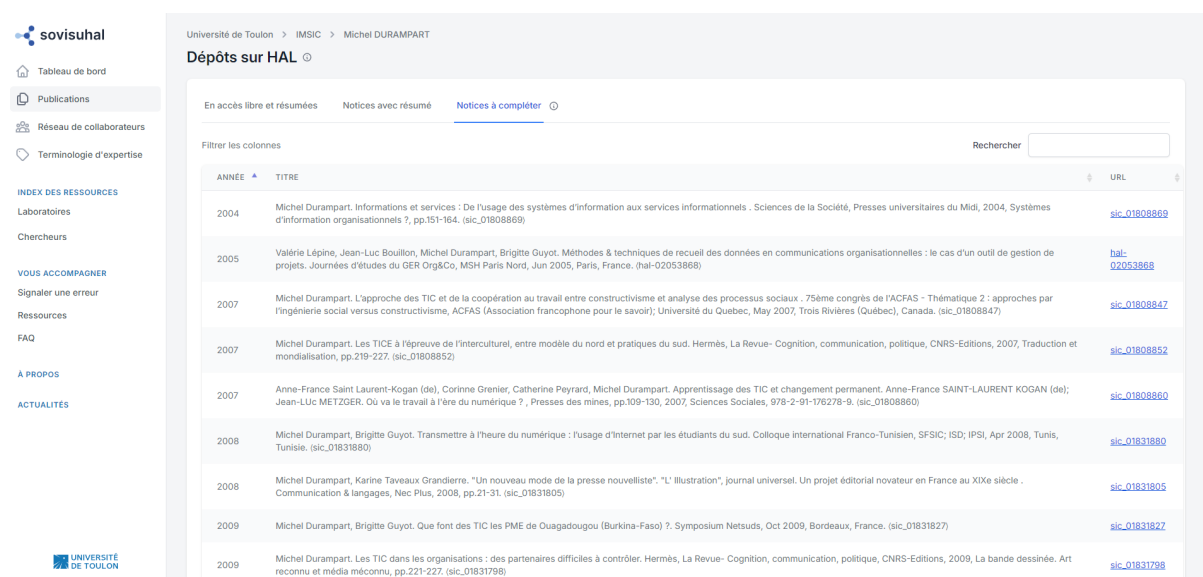
Capture d'écran du tableau de bord du premier prototype



Une seconde vue (figure 7) détaillée décline la liste des notices moissonnées rattachées à l'auteur. Cette vue présente chaque publication avec ses métadonnées associées (titre, date de publication, auteurs, etc.) et décline l'affichage selon des indicateurs spécifiques, comme la présence ou l'absence de résumés ou d'autres éléments clés dans les dépôts. Cette présentation facilite l'identification rapide des publications complètes et surtout incomplètes pour l'utilisateur afin que l'utilisateur puisse compléter les dépôts qui le nécessite.

Figure 7

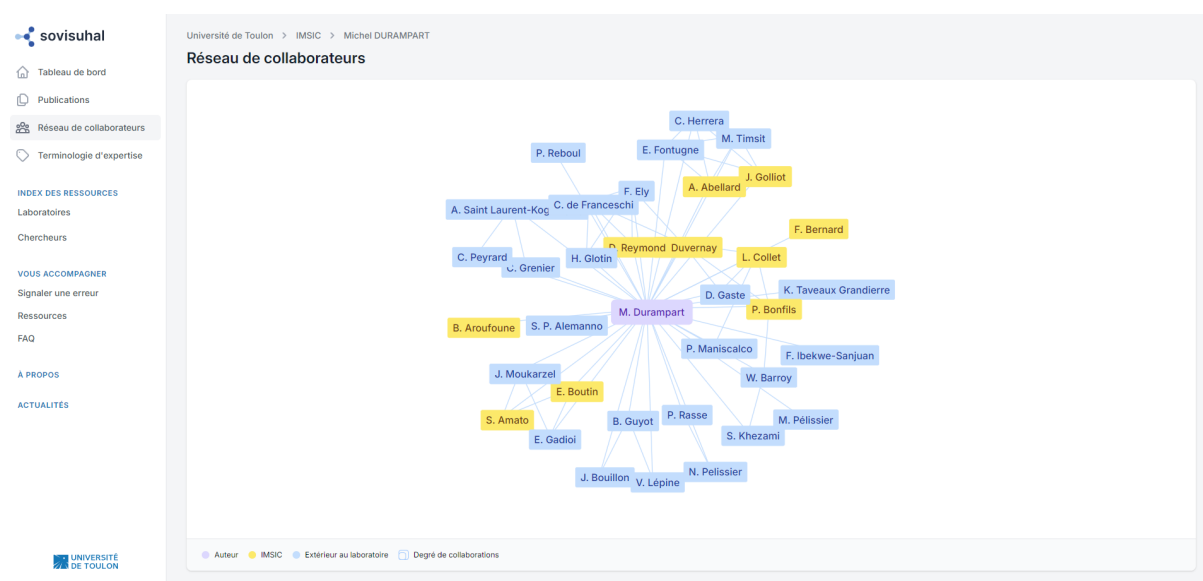
Capture d'écran de la liste des publications catégorisées du premier prototype



Une troisième vue (figure 8) visualise un réseau de collaboration basé sur les co-auteurs identifiés dans les notices d'un auteur donné. Le réseau distingue clairement les chercheurs internes à l'entité d'appartenance de l'auteur des collaborateurs externes, ce qui facilite une lecture rapide des dynamiques collaboratives internes et externes, autour d'un chercheur, au sein du laboratoire ou de l'université.

Figure 8

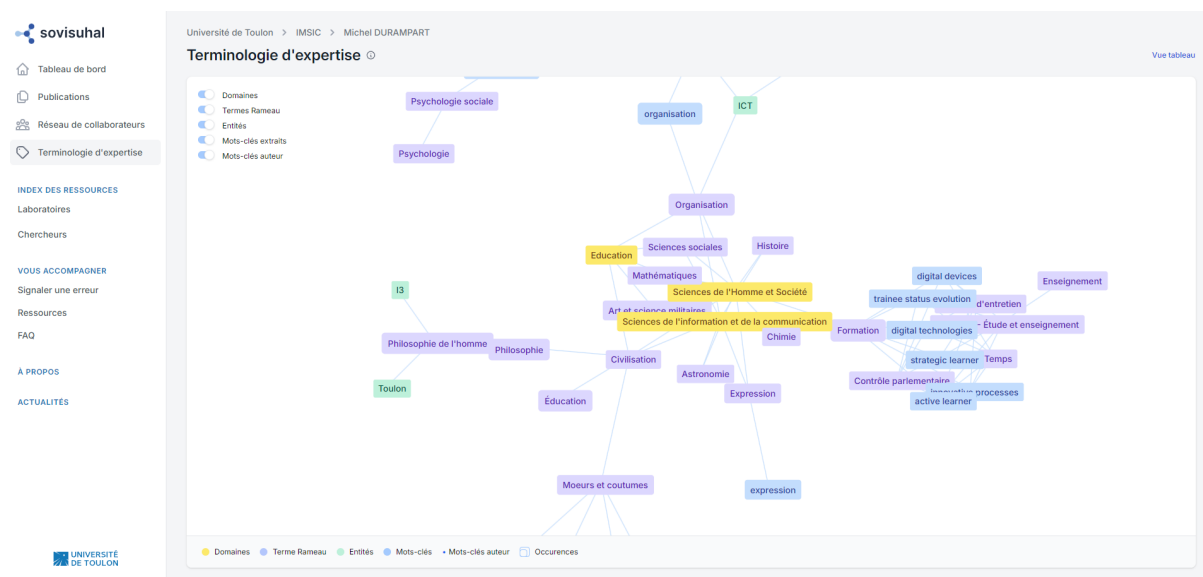
Capture d'écran de l'exploration du réseau de collaboration du premier prototype



Enfin, une quatrième vue (figure 9) se concentre sur l'analyse des termes et concepts issus des métadonnées des dépôts. Elle intègre à la fois les mots-clés directement associés aux publications, les mots-clés extraits automatiquement, et les entités identifiées dans les métadonnées. En complément, un enrichissement est effectué grâce à l'utilisation du vocabulaire contrôlé issu de RAMEAU (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié), le langage d'indexation de référence en France (Freschard, 1990). Cette vue permet une exploration approfondie des thématiques abordées et une meilleure compréhension des orientations scientifiques de l'auteur, du laboratoire ou encore de l'université.

Figure 9

Capture d'écran de l'exploration terminologique du premier prototype



Les résultats de cette première expérimentation s'avèrent prometteurs, mais ils mettent également en lumière certaines limites. Bien que la visualisation soit particulièrement éloquent sur le plan représentationnel, notamment en termes de synthèse visuelle et cartographique, plusieurs aspects soulèvent des questionnements. D'une part, les cartes générées, réalisées sans la participation active des chercheurs, risquent de ne pas refléter de manière exhaustive la réalité de leurs productions. Elles peuvent manquer d'informations cruciales ou présenter des biais liés à l'incomplétude des dépôts. D'autre part, la réduction de la notion d'expertise à ce qui est déposé dans HAL et décrit par le chercheur constitue une simplification qui peut occulter d'autres facettes de la compétence scientifique. En d'autres termes, bien que les représentations synthétiques des productions soient visuellement efficaces, dynamiques et interactives, elles peuvent s'avérer erronées si les données ne sont pas vérifiées, rejoignant ainsi les nombreuses mises en garde de travaux du domaine des systèmes d'information sur la recherche (en anglais, *Current Research Information System, CRIS* ; Bryant et al., 2021 ; R et al., 2019). Ce constat est également en résonance avec les réflexions développées par Marty et Gallezot (2020), qui rappellent que l'analyse et la visualisation de données ne sont jamais neutres, mais s'inscrivent dans un cadre d'interprétation nécessitant des précautions méthodologiques et épistémologiques. Selon eux, la visualisation peut certes constituer un puissant outil heuristique pour la recherche en SIC, mais elle doit être pensée comme un processus de médiation, à la fois technique, documentaire et symbolique, entre les données et leur mise en sens.

Face à ces limites, nous proposons un dispositif d'accompagnement pour une science ouverte dans une logique mutuellement avantageuse. D'une part, ce dispositif incite le chercheur à auto-archiver ses productions, tout en lui permettant de vérifier et compléter ses données. En retour, ces données vérifiées puis probablement enrichies contribuent à une meilleure indexation de son profil et participent à une représentation plus fidèle et exhaustive de son expertise. Une telle approche répond implicitement aux demandes de recensement exhaustif demandées dans le cadre des campagnes d'évaluation du Hcéres et vise également à améliorer la complétude des données, renforçant ainsi la qualité des visualisations générées.

2.3. Extension du développement dans une logique d'accompagnement

Le dispositif s'est enrichi progressivement d'une logique d'accompagnement structurée, visant à dépasser le simple affichage de données issues de HAL. Dans cette perspective, SoVisu peut être lu comme un outil de gouvernementalité documentaire (Chartron, 2010), qui ne se contente pas de visualiser des données mais oriente les usages, rend visibles certains types d'informations et, ce faisant, participe à la normalisation des pratiques de dépôt. Il s'agit donc d'un instrument de médiation, mais aussi de prescription : ce que SoVisu rend visible devient potentiellement ce qui compte, ce qui est compté. Il a été conçu pour favoriser un engagement cognitif et informationnel des chercheurs via une interface lisible, actionnable et réversible. La visualisation n'est donc pas ici pensée comme un outil de reporting, mais comme un levier de médiation (Couzinet & Liquète, 2021). Dans cette perspective, nous mobilisons les travaux de Schmitt (2018), qui propose un paradigme méthodologique articulé autour de l'exploration, de la visualisation et de la décision à partir de grandes masses de données hétérogènes. Schmitt montre que les dispositifs numériques doivent être conçus non seulement pour organiser et traiter l'information, mais aussi pour rendre les connaissances produites intelligibles, interprétables et susceptibles de guider l'action des utilisateurs. Cette approche éclaire directement la conception du dispositif SoVisu, qui vise à étendre l'expérience d'archivage scientifique sur HAL en apportant une couche d'exploration visuelle et de médiation documentaire fondée sur les métadonnées, sans se substituer aux fonctionnalités natives de l'archive. À l'instar de Schmitt, nous adoptons une démarche réflexive intégrant à la fois des méthodes quantitatives et qualitatives, dans une perspective où la technique est pensée comme un levier pour renforcer l'engagement des chercheurs dans la science ouverte.

D'un point de vue technique, la seconde version du prototype est développée en utilisant le langage de programmation Python. Les développements commentés ci-après sont publiés et librement accessibles sur HAL (Reymond et al., 2022b). Pour préserver la lisibilité de cette section, nous choisissons d'éviter les détails techniques. Le lecteur curieux est invité à consulter le dépôt, où ces éléments ainsi que la documentation propre au dispositif (Reymond et al., 2022a) sont accessibles. Ils se décomposent en deux parties modulaires, à savoir, une partie dédiée à la collecte et aux traitements de l'information et une seconde partie dédiée à l'interface et proposant des visualisations analytiques accessibles à l'utilisateur.

2.3.1. Initialisation du dispositif

Le dispositif SoVisu est initialisé à partir de plusieurs fichiers texte au format CSV, constituant la structure de l'organisation. Le premier fichier correspond aux établissements de tutelle. Il comporte l'identifiant sur le Système d'identification du répertoire des entreprises (SIREN), le nom complet ainsi qu'un acronyme pour chaque établissement. Le second fichier détaille le niveau intermédiaire, constitué par les laboratoires de recherche. Celui-ci inclut plusieurs identifiants comme l'identifiant de la structure sur l'archive HAL, l'identifiant dans le Répertoire National des Structures de Recherche (RNSR) et celui sur le référentiel Identifiants et Référentiels pour l'Enseignement Supérieur et la Recherche (IdRef). À ces données s'ajoutent le nom, l'acronyme du laboratoire et l'identifiant de l'établissement parent, assurant ainsi la structuration hiérarchique entre les niveaux organisationnels. Le troisième fichier traite du niveau individuel, c'est-à-dire les chercheurs. Il liste les structures d'appartenance des chercheurs, leurs différents identifiants internes (identifiant issu de l'annuaire institutionnel) et externes avec l'identifiant sur l'archive HAL et sur le référentiel IdRef, ainsi que les informations issues des champs définis par la norme pour les Annuaire pour l'Enseignement Supérieur (Supann). Ainsi, l'ensemble des fichiers combine des données d'identification institutionnelles internes et des données externes nécessaires à l'intégration dans l'écosystème de la science ouverte. Seulement quelques lignes d'informations suffisent pour initialiser le dispositif, ces données étant automatiquement complétées par le processus d'authentification auquel l'utilisateur a accès et qui est décrit dans les sections suivantes.

2.3.2. La collecte et le traitement de l'information

Avant de traiter la collecte et l'exploitation des informations, il est essentiel de décrire le point d'entrée par lequel le chercheur s'enregistre sur la plateforme. Nous détaillons ici la phase

préalable à la collecte. L'interface utilisateur qui guide ces premiers pas et les options qu'elle propose sont abordées plus en détail dans la section suivante.

Une entrée via les chercheurs

Le chercheur s'authentifie par l'intermédiaire du Central Authentication Service (CAS ; Mathieu et al., 2003). Pour l'utilisateur, il s'agit du portail de connexion proposé au chercheur lorsqu'il souhaite accéder aux services numériques d'une université. Par l'intermédiaire de l'annuaire LDAP, cette authentification génère une identification des usagers en récupérant différentes données comme le nom, le prénom ou encore l'affiliation, pour réaliser un assemblage hiérarchique simple et suffisant dans le cadre de l'Université de Toulon : le chercheur est affilié à un laboratoire, lui-même associé à l'établissement. Cet assemblage hiérarchique est essentiel pour structurer les données et garantir une identification précise des chercheurs et de leurs affiliations, facilitant ainsi l'analyse des collaborations et de la production scientifique. Les péripéties du projet SoVisu nous montrent qu'un tel assemblage est difficilement transposable à des universités plus « complexes », comme celle de Paris-Saclay. Un premier processus automatisé d'association des données d'identification dans l'écosystème de la science ouverte se déclenche en arrière-plan. Nous moissonnons diverses plateformes, dont HAL, à l'aide de scripts dédiés, afin de récupérer les identifiants associés au nom du chercheur : IDRef, ORCID et identifiant HAL. Les identifiants retrouvés sont proposés par une interface à l'utilisateur, qui peut alors les valider, les compléter ou les corriger. Si l'identifiant HAL d'un chercheur n'est pas déterminé, l'utilisateur est invité à se créer son identifiant et à le rapporter dans le dispositif.

Une fois ce processus achevé, notre système moissonne l'ensemble des métadonnées des dépôts HAL liés à l'identifiant validé en interrogeant l'interface de programmation mise à disposition, ainsi que les données structurées associées à l'entité, directement depuis dataHAL. Des informations complémentaires sont ensuite calculées ou récupérées afin d'affiner la caractérisation des dépôts. Ces traitements détaillés ci-après visent à enrichir la lecture des métadonnées HAL et à renforcer l'analyse des pratiques d'archivage.

Nous intégrons des services externes pour enrichir l'analyse des métadonnées des dépôts. Par exemple, nous avons recours à Teeft, un outil d'analyse lexicale développé par l'INIST-CNRS, qui permet l'extraction automatique de mots-clés à partir des résumés des dépôts (Cuxac et al., 2019). Nous interrogeons également l'API Unpaywall pour calculer des métriques relatives à l'accès libre des documents, comme leur type, leur statut ou encore l'entrepôt où se trouve la ressource en question, à partir de leur identifiant DOI (Piwowar et al., 2018). Cette démarche repose toutefois

sur la disponibilité effective du DOI dans les métadonnées HAL, ce qui constitue une limite. Certaines publications, notamment issues de conférences ou de champs moins bien normalisés, peuvent ne pas disposer de DOI, ou ne pas l'avoir renseigné dans l'archive.

Une mesure de qualité des dépôts

Par ailleurs, nous vérifions systématiquement la présence de champs descriptifs essentiels (fichiers, mots-clés, résumé), indépendamment du langage, et nous attribuons une métrique évaluant le niveau de description de la ressource concernée (Tabariès, 2022). Cette approche s'inscrit dans les réflexions de Jean Meyriat sur le rôle structurant des systèmes documentaires dans l'organisation du savoir scientifique. Par extension logique, cette approche s'aligne également avec les principes FAIR (Wilkinson et al., 2016), qui soulignent l'importance de métadonnées complètes et normalisées pour assurer la découvrabilité et l'exploitation des ressources scientifiques. En effet, comme déjà évoqué auparavant, la richesse et la précision des métadonnées ne se limitent pas à un enjeu technique : elles déterminent les modalités d'organisation, d'accès et de réutilisation de l'information scientifique. Par souci de praticité, nous nous concentrons sur la complétude formelle des notices sans évaluer la qualité des informations renseignées. Toutefois, pour éviter les biais liés aux résumés trop succincts, nous considérons comme absents ceux dont la longueur est inférieure ou égale à celle du titre, ce qui permet également d'exclure les entrées telles que « no abstract ». Notre « score de découvrabilité », désigné dans le dispositif sous l'intitulé « QD » pour qualité de description, dans l'interface, reflète ainsi le degré de structuration des dépôts et leur potentiel de découvrabilité et, par extension, d'exploitation pour la production de nouveaux savoirs. Dans l'article cité, les différentes métadonnées contribuant à cette métrique étaient initialement pondérées de manière arbitraire. Nous privilégions désormais une pondération uniforme afin d'assurer une évaluation plus cohérente du niveau de découvrabilité des ressources. Cette approche s'inscrit dans une démarche d'homogénéisation des pratiques d'évaluation, rejoignant ainsi les conclusions de Gueguen (2019), qui souligne l'importance d'une approche standardisée pour garantir une évaluation plus robuste de la qualité des métadonnées à grande échelle.

Des travaux plus récents pourraient permettre de faire évoluer à nouveau la métrique, ce qui n'est pas fait ici. La découvrabilité des travaux scientifiques dépend fortement de la qualité, de la structuration et du multilinguisme des métadonnées, notamment les titres, résumés et mots-clés (COAR Task Force, 2023 ; Francoeur et al., 2025). Plusieurs études internationales démontrent que fournir ces éléments dans différentes langues améliore significativement l'accessibilité et la visibilité des publications, en permettant leur repérage par des publics diversifiés et dans des

contextes non anglophones (Di Bitetti & Ferreras, 2017 ; Ramírez-Castañeda, 2020). Dans cette perspective, l'amélioration de la qualité descriptive multilingue constitue un levier essentiel pour renforcer l'impact et la circulation des savoirs dans les environnements numériques contemporains et, en conséquence, peut être prise en compte dans le calcul de la métrique présentée précédemment.

Indexation

Les informations moissonnées sont téléversées vers une instance Elasticsearch (*elastic/elasticsearch*, 2010/2024), qui sert de base de données et de moteur d'indexation. Sur le plan technique, et en lien avec la technologie utilisée, les index, qui intègrent les données, sont générés pour chaque chercheur selon une structure normalisée combinant plusieurs identifiants, à savoir, celui de l'université, du laboratoire, l'identifiant LDAP du chercheur, ainsi qu'une mention explicite du type d'entité en question et de la nature de l'information. Cette structuration permet de concevoir des vues de données facilitant l'interrogation d'ensembles spécifiques, tels que les chercheurs affiliés à un laboratoire ou à une université. Une fois la collecte automatisée finalisée, le chercheur reprend la main pour valider les informations qui le concernent. Cette étape marque la transition entre une extraction brute des données et une qualification par l'utilisateur. Celui-ci peut alors exclure certains dépôts associés à son profil s'il estime, par exemple, que son implication y a été mineure, affiner les métadonnées le décrivant, en précisant ses domaines d'expertise selon la nomenclature utilisée sur l'archive et ajouter des informations complémentaires le concernant. Ce point est détaillé plus loin dans le commentaire de l'interface.

Nous présentons ici un processus que l'on peut qualifier d'automatisation guidée, visant à améliorer la collecte de données en comblant certaines lacunes du premier prototype. Ce modèle repose sur une interaction entre automatisation et intervention humaine, remplaçant le chercheur au centre du processus de qualification et de validation des données qui le concernent. Cette approche s'inscrit dans le concept de « document par intention » développé par Jean Meyriat (1981), qui souligne que l'information ne se limite pas à sa simple existence mais dépend de son contexte d'usage et de sa réinterprétation par les utilisateurs. Dans le cadre de notre dispositif, cela signifie que le système ne se base pas sur l'information auparavant archivée, mais laisse la possibilité au chercheur de requalifier l'information en validant, enrichissant ou corrigeant les métadonnées qui lui sont associées. L'intervention du chercheur transforme un simple dépôt en un document structuré et exploitable dans les systèmes d'information scientifique. Cette intervention peut correspondre à une première qualification, lorsque les dépôts en question sont déposés par un intermédiaire (co-auteur, bibliothécaire, etc.), ou à une nouvelle qualification où le

contexte peut être réévalué, lorsqu'il s'agit de valider, compléter ou d'exclure de son profil un dépôt déposé en auto-archivage.

L'implication directe du chercheur dans la qualification et la valorisation de l'information scientifique ne se limite pas à une simple correction technique : il joue un rôle actif de médiateur de ses propres données. Cette démarche rejoint les travaux de Couzinet et Liquète (2021) sur la médiation documentaire, qui soulignent l'importance de l'intervention humaine dans la mise en circulation des savoirs. En s'appropriant ce processus, le chercheur ne se contente plus d'utiliser HAL comme un outil de dépôt unitaire : il devient acteur de la structuration de ses données, en validant, corrigeant ou enrichissant les métadonnées associées à ses productions. C'est précisément ce que permet le dispositif SoVisu, en complément de HAL : alors que ce dernier repose sur une logique formulaire linéaire, SoVisu propose une interface de synthèse et de visualisation qui rend visibles les incohérences ou manques dans les notices, et qui invite à une reprise critique et contextualisée de ces métadonnées, qui doit cependant être effectuée dans l'archive HAL elle-même, seule habilitée à enregistrer les modifications. Autrement dit, SoVisu comble une lacune de HAL en replaçant le chercheur dans une posture active et réflexive sur l'ensemble de sa trajectoire documentaire. Ce dernier peut affiner ses domaines d'expertise, compléter les notices incomplètes, ou encore désassocier certaines productions qui ne reflètent pas fidèlement son activité scientifique. Ce rôle élargi, qui va au-delà du simple dépôt, renforce la responsabilisation du chercheur quant à la qualité de son profil dans l'écosystème de la science ouverte. Il ouvre également la voie à des formes de valorisation plus qualitatives, en cohérence avec les approches émergentes d'évaluation narrative (Boukacem-Zeghmouri, 2023), où le sens donné aux productions et leur mise en récit comptent autant que leur nombre ou leur visibilité brute.

2.3.3. *La question des laboratoires*

Par construction et en rapport avec les données de l'annuaire, à quelques exceptions individuelles près que nous ignorons par la suite, les affiliations des chercheurs permettent d'associer leurs données à un laboratoire. Le regard sur la production de ces entités, peut s'effectuer selon deux perspectives : d'une part, à travers la somme de la production scientifique des chercheurs actuellement affiliés au laboratoire, ce qui à notre connaissance est rare pour un dispositif et, en même temps, peut se réaliser sur les rapports Hcéres, et d'autre part, en considérant l'ensemble des productions des chercheurs ayant été ou étant encore affiliés via les plus traditionnelles « collections de laboratoire » constituées de notices tamponnées par les gestionnaires de collection, sur l'archive, en fonction des finalités retenues. Ainsi, lors de l'initialisation d'un laboratoire,

processus automatisé, les métadonnées des dépôts associées à son identifiant HAL sont moissonnées, puis téléversées vers l'instance Elasticsearch. Une fois encore, les index sont structurés à partir des identifiants afin de faciliter l'interrogation et l'agrégation des ensembles de données. Les deux perspectives mentionnées peuvent donc être explorées indépendamment par la suite, en fonction des objectifs envisagés.

Notons que les deux perspectives évoquées ouvrent la voie à des approches différenciées de l'indexation et de la structuration des connaissances au sein des laboratoires. La première perspective, ancrée dans une démarche historique, suppose une reconstitution dynamique de l'évolution du laboratoire à travers l'analyse de ses productions scientifiques et archivistiques. Cette approche cherche à comprendre comment les savoirs se sont constitués, transformés ou consolidés dans le temps, en mobilisant les archives institutionnelles et les dépôts associés comme matériaux de recherche. Chaque notice est alors appréhendée comme une trace épistémologique, révélatrice des évolutions successives des cadres théoriques, méthodologiques et institutionnels du laboratoire sur plusieurs décennies. Elle rejoint ainsi les réflexions de Kembellec et al. (2017) sur la mise en archive de l'histoire scientifique et la structuration progressive des savoirs scientifiques à travers un dispositif archivistique en sciences de l'information et de la communication. Les auteurs montrent notamment comment un système d'archivage numérique permet de croiser différentes typologies de documents et d'acteurs pour analyser les dynamiques de construction du savoir scientifique dans un cadre institutionnel donné, c'est-à-dire les processus concrets, souvent évolutifs, par lesquels les connaissances sont produites, validées, diffusées et archivées au sein d'un environnement organisationnel et documentaire.. À l'inverse, la seconde perspective repose sur une genèse en devenir, une cartographie dynamique de la production scientifique actualisée en fonction des chercheurs actuellement affiliés au laboratoire. Il s'agit d'une vision prospective, orientée vers le potentiel de genèse, c'est-à-dire la capacité du laboratoire à engendrer de nouveaux savoirs en fonction de la structuration actuelle de son réseau de recherche. De Biasi (2003) met en évidence la tension entre la trace archivistique et la dynamique de production des savoirs, en adaptant la génétique des textes aux archives scientifiques. Il distingue ainsi un stade génésique, où l'écriture accompagne la recherche dans une phase exploratoire, et un stade textuel, où elle structure et communique les résultats. Cette distinction éclaire la complémentarité entre une indexation rétrospective des savoirs et une cartographie dynamique des productions en cours. Si la première approche s'apparente à une stratigraphie épistémologique, permettant de reconstruire l'évolution des savoirs selon une logique rétrospective et cumulative, la seconde vise à rendre compte d'un processus émergent. Dans cette perspective, la question de la structuration documentaire prend une ampleur

stratégique : faut-il privilégier une indexation diachronique, fidèle à la trajectoire historique du laboratoire, ou une indexation synchronique, focalisée sur les productions des chercheurs actuellement associés au laboratoire ? Le modèle d'assemblage documentaire envisagé détermine ainsi la nature des visualisations produites : dans une approche historique, les indicateurs révèlent la généalogie des concepts et des méthodologies, tandis que dans une approche dynamique, ils permettent d'anticiper les trajectoires possibles d'innovation scientifique. Ces deux modèles ne s'opposent pas mais se complètent, offrant des cadres de représentation différenciés selon les objectifs poursuivis. L'adoption d'un dispositif hybride, combinant à la fois une historisation des productions scientifiques et une analyse en temps réel des dynamiques de recherche, apparaît comme une voie pertinente pour articuler les deux logiques. Une telle approche permettrait de dépasser l'opposition entre construction historique et potentiel de genèse, en intégrant dans un même référentiel à la fois les traces du passé scientifique du laboratoire et ses potentialités futures.

2.3.4. L'interface utilisateur

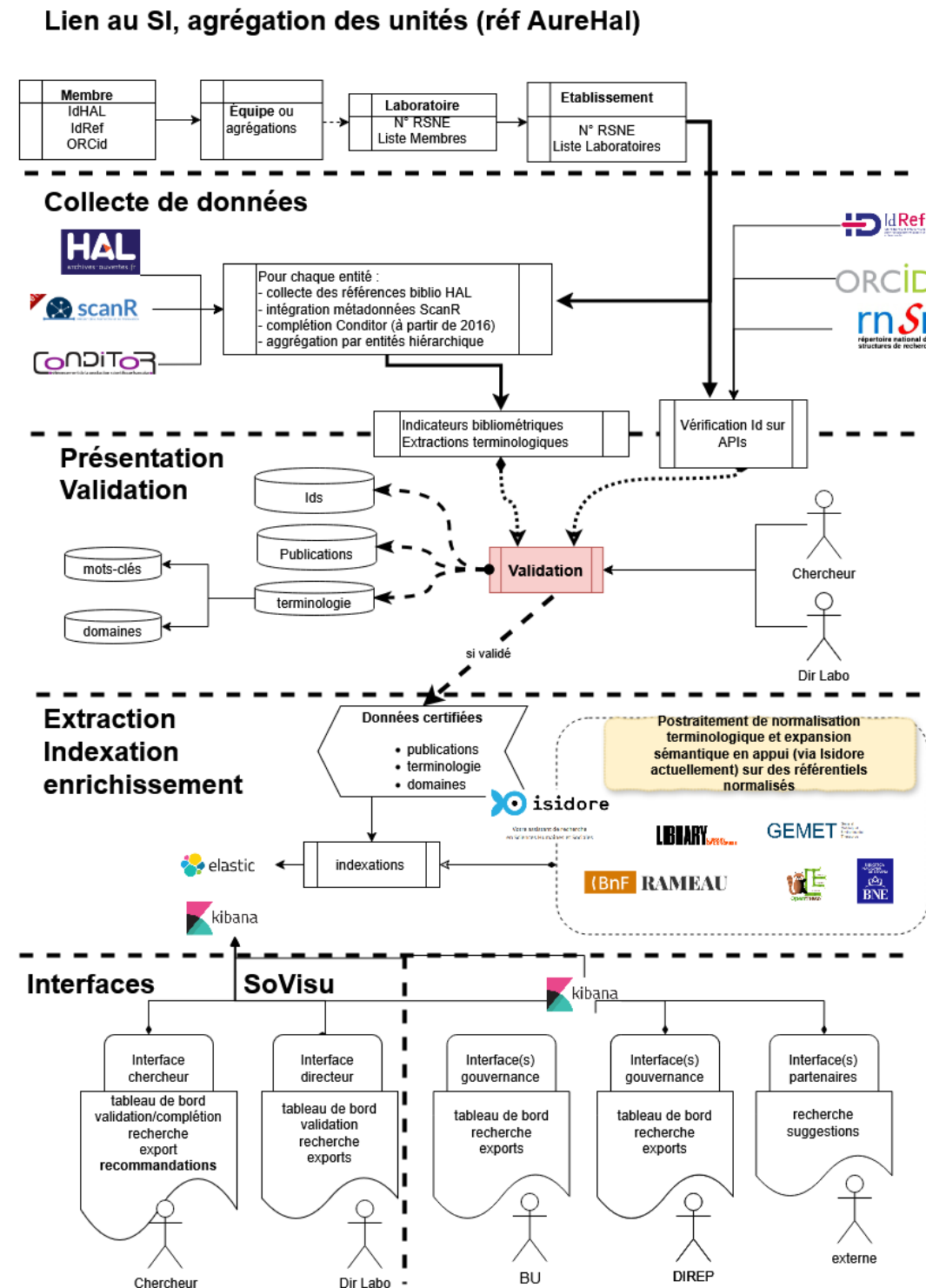
L'interface du dispositif adopte une structure proche d'une application web, privilégiant une navigation fluide et un accès direct aux contenus. Elle se divise en trois sections principales. Sur la gauche, un menu latéral propose plusieurs catégories facilitant l'exploration des données. La première, « profil », regroupe les informations liées à l'utilisateur, qu'il peut compléter ou modifier. La seconde, « synthèses », rassemble des visualisations offrant une vue d'ensemble des données associées au chercheur. La catégorie « ressources » donne accès aux entités créées et caractérisées, comme les laboratoires et les chercheurs. « Exploration » met à disposition des outils permettant d'interroger le système d'information selon différentes finalités. Enfin, « plus d'informations » réunit des ressources internes et externes pour mieux comprendre le dispositif et son écosystème. L'interface privilégie une organisation centrée sur le contenu. L'espace principal affiche directement les informations de la page sélectionnée, réduisant ainsi la friction entre navigation et consultation. Lorsque cela s'avère nécessaire, les informations de la page peuvent être segmentées en onglets pour faciliter leur consultation. Cette structuration est renforcée par un titre en haut de l'écran, guidant l'utilisateur dans son parcours. À la droite du titre, un sélecteur discret permet à l'utilisateur de choisir une langue ainsi qu'une période pour affiner les données sur lesquelles reposent les affichages. Ces choix de conception reposent sur plusieurs principes fondamentaux visant à optimiser l'expérience utilisateur et l'efficacité du dispositif. Tout d'abord, une navigation intuitive permet de minimiser la charge cognitive, en facilitant l'accès aux informations clés. L'approche modulaire et interactive, inspirée des applications web, assure une

expérience fluide et engageante, s'éloignant d'un modèle statique traditionnel. La segmentation des contenus en catégories distinctes répond également aux besoins spécifiques des chercheurs et des personnels impliqués dans la recherche, en mettant à leur disposition des outils adaptés à leur activité. Enfin, cette organisation vise à structurer l'information de manière efficace, en garantissant une hiérarchisation claire et une meilleure lisibilité des contenus. Nous présentons par la suite l'ensemble des pages de l'application en suivant le parcours naturel de l'utilisateur, afin d'illustrer concrètement ces choix de conception et leur impact sur l'interaction avec le dispositif.

Sur le plan technique, les visualisations présentées ci-après sont générées et affichées à l'aide de Kibana (*elastic/kibana*, 2013/2024), un dispositif complémentaire à Elasticsearch, précédemment mentionné. Kibana permet de transformer des ensembles de données volumineux en représentations graphiques exploitables, facilitant ainsi l'exploration et l'analyse des informations. Grâce à ses capacités avancées en matière de tableaux de bord interactifs, filtres dynamiques et visualisations personnalisées, il offre aux utilisateurs une vue synthétique et structurée des données. Cette intégration optimise l'accessibilité des informations et contribue à une meilleure compréhension des tendances et des corrélations au sein du corpus analysé. Il est intéressant de noter que Kibana est désormais utilisé comme module de statistique pour les données archivées sur l'archive ouverte HAL (*Kibana – com&doc*, s. d.).

Figure 10

Schéma fonctionnel du dispositif SoVisu



Lors de sa première connexion, le point d'entrée du dispositif SoVisu permet à l'utilisateur d'initialiser son profil chercheur. L'utilisateur s'authentifie via le portail CAS de son institution. Une fois authentifié, le dispositif lui propose de remplir un formulaire (figure 10), pré-rempli par

un processus automatisé avec des informations suggérées, obtenues par des requêtes à des services externes, ainsi que par l'interrogation de l'annuaire LDAP associé à l'utilisateur au niveau même de l'institution. Ce formulaire présente à l'utilisateur plusieurs identifiants clés dans l'écosystème de la recherche, à savoir les identifiants HAL, ORCID et IDRef. Les deux premiers sont les plus utilisés par les chercheurs en France, représentant respectivement les dimensions nationale et internationale (Bouchard & Boudry, 2024). IDRef, quant à lui, est principalement utilisé pour l'identification normalisée des chercheurs et des autorités dans les bases documentaires et bibliographiques. Le formulaire inclut également des informations de base sur l'utilisateur, comme son statut et sa structure. La validation par l'utilisateur de ces informations garantit un moissonnage et un classement précis des données liées à l'utilisateur dans le système. Une fois ces données validées, le processus de collecte décrit précédemment est initié. Après quelques secondes d'attente, les autres pages du dispositif s'offrent à l'utilisateur. Nous proposons au lecteur de les découvrir ci-après.

Profil

La section « profil » comprend un unique lien permettant à l'utilisateur d'afficher et de vérifier les données moissonnées ou précédemment vérifiées le concernant (figure 11). Pour rappel, par l'expression « vérifier », nous entendons la possibilité de valider, compléter ou retirer certaines informations de son profil. Ainsi, l'utilisateur peut consulter et gérer ses identifiants associés, les notices HAL moissonnées et liées à son identifiant HAL, distinguées en notices validées et notices retirées, ainsi que son profil d'expertise moissonné et rattaché à son identifiant HAL. Il peut également vérifier les informations relatives à l'autorat de chaque notice (par exemple, s'il est premier auteur ou auteur correspondant). L'utilisateur a aussi accès à l'extension de son domaine d'expertise, représentée sous forme d'arbre structuré suivant la nomenclature présente sur HAL. En complément, un dernier onglet dédié aux descriptifs de recherche apporte une dimension qualitative supplémentaire à son profil. Cet espace permet de renseigner des mots-clés généraux, un résumé des travaux en cours, ainsi que des informations sur les projets en cours et leurs financements. De par cette organisation, l'utilisateur peut ainsi contrôler et ajuster son profil de manière exhaustive, garantissant la cohérence et la pertinence des informations qui lui sont associées.

Figure 11

Capture d'écran de la page « vérifier ses données » du dispositif SoVisu

Vérification des données > A. TABARIES

FENÊTRE DE DÉPÔT
1990 - 2025

PROFIL
Vérification des données

SYNTHÈSES
Tableau de bord
Références
Expertise
Lexiques extraits

RESSOURCES
Index des laboratoires
Index des chercheurs

EXPLORATION
Recherche d'expert

PLUS D'INFORMATIONS
Lire la documentation
Présentation
Ressources
Foire aux questions
Déconnexion

Identifiants Notices HAL Expertises identifiées Domaines d'expertise

Dans cet onglet, vous devez vérifier, renseigner ou corriger vos identifiants chercheur.
L'**IdHal** est votre identifiant sur l'archive nationale Hal. Il permet de regrouper vos publications et de gérer les différentes formes auteur. Cette fonction est à réaliser via HAL.
L'**IdRef** s'obtient dès la publication de votre thèse. Vous pouvez le vérifier : [sur le site prévu à cet effet](#)
L'**Orcid** est un identifiant chercheur international. Nous vous recommandons d'en créer un. Vous avez une petite procédure réalisée par l'UMR 5206 [sur ce lien](#).
Si vous avez d'autres identifiants (Arxiv, Mendeley, Research Gate...) leurs moissonneurs se chargeront de publier automatiquement vos notices sur ou depuis HAL.
Merci de vérifier ces identifiants avant de les renseigner ici, puis validez. Vous accédez alors aux autres fonctionnalités de SoVisu !

ID HAL (texte)
alaric-tabaries

aurehal ID
12246113

laboratoire affilié
IMISIC

Statut
Doctorant

IdRef
-

ORCID
-

Valider les identifiants

La page permettant de vérifier les données associées au profil d'un utilisateur est également adaptée pour les laboratoires. Dans ce cas, elle permet d'afficher la liste des chercheurs affiliés à un laboratoire et d'y associer un numéro d'équipe spécifique, offrant ainsi une granularité plus fine pour d'éventuels traitements futurs. La structure et les finalités des autres onglets est similaire à celle du profil utilisateur, garantissant une organisation cohérente des informations entre les différents niveaux d'analyse.

Synthèses

La section « synthèses » constitue un élément central du dispositif SoVisu. Elle offre à l'utilisateur un accès structuré à différentes ressources analytiques liées à son profil et à ses dépôts validés. Parmi ces ressources figurent un tableau de bord généraliste permettant de synthétiser et de visualiser ses données, un accès à la liste de ses dépôts triés selon divers critères, ainsi qu'une vue détaillée de ses expertises. Un autre tableau de bord plus spécifique est également disponible pour analyser le lexique extrait des métadonnées décrivant ses dépôts.

Le tableau de bord interactif (figure 12), conçu pour offrir une vision globale des contributions de l'utilisateur, présente plusieurs indicateurs clés facilitant l'analyse de son activité sur HAL. Ce tableau de bord présente plusieurs indicateurs clés. Il affiche d'abord le nombre total de dépôts, permettant ainsi de quantifier la contribution scientifique de l'utilisateur sur l'archive, ainsi que le

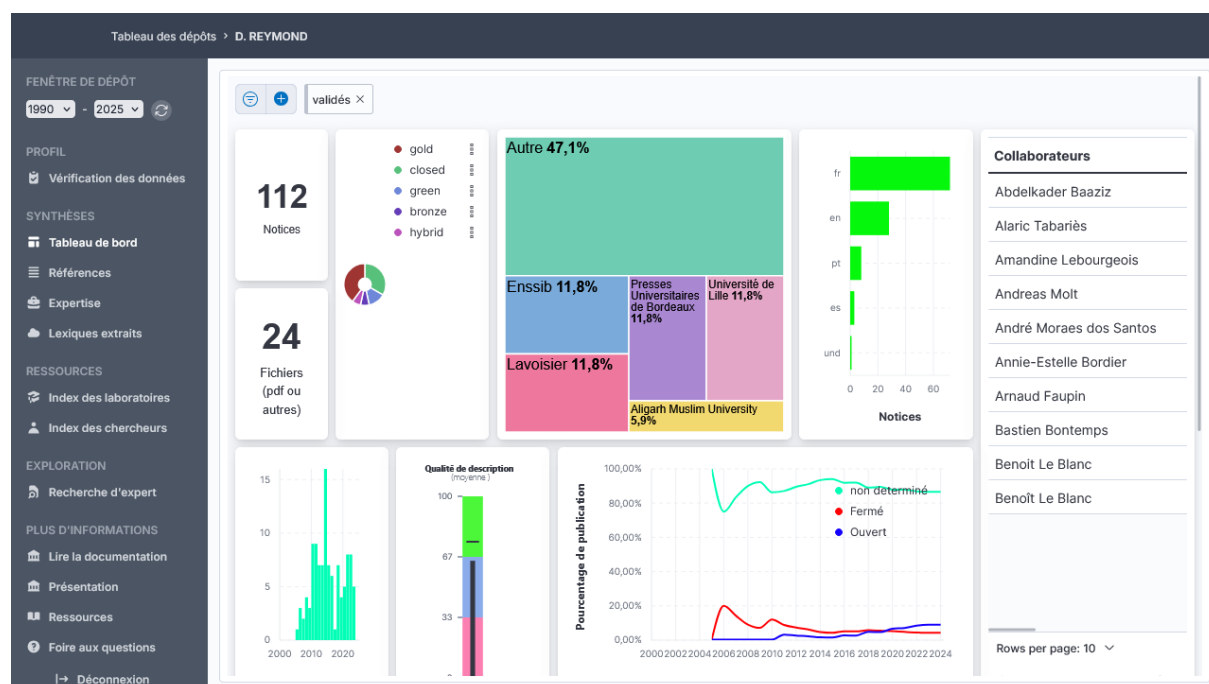
nombre de dépôts contenant un fichier associé, souvent le texte intégral. Cette distinction est particulièrement intéressante, car, au-delà de mesurer les dépôts offrant une valeur ajoutée en termes de diffusion de l'information scientifique, elle s'inscrit dans un contexte plus large où la science ouverte devient un enjeu stratégique. Plusieurs institutions accordent désormais une importance spécifique à ces données dans le cadre des politiques d'évaluation et des dispositifs d'incitation, notamment via des indicateurs de type « bonus performance ». La volumétrie des dépôts dans le temps est représentée sous forme d'histogramme, mettant en évidence les tendances et les variations au fil du temps. La diversité linguistique des publications est également prise en compte à travers une analyse de la répartition des langues des notices déposées. Un autre élément clé du tableau de bord est l'analyse des voies de publication, qu'elles soient ouvertes, fermées ou indéterminées, avec une représentation en pourcentages pour faciliter la comparaison. La répartition des éditeurs publiant les travaux déposés est quant à elle visualisée sous forme de carte proportionnelle, permettant d'identifier les maisons d'édition les plus représentées. L'évolution du nombre de publications par année et par type d'accès est également mise en évidence. Une courbe de tendance en valeurs absolues et en pourcentages permet de mieux appréhender les pratiques de dépôt sur le long terme. Enfin, un tableau récapitulatif des collaborateurs les plus fréquents, basé sur l'analyse des co-auteurs des notices déposées, apporte un éclairage sur les dynamiques de collaboration scientifique. Les données utilisées pour la génération des visualisations peuvent être déclinées selon le type de travail (article, communications, etc.) en utilisant les sélecteurs sur la partie haute.

Toutefois, si ce tableau de bord offre une vue d'ensemble riche et détaillée, il peut également sembler brouillon ou fragmenté dans la manière dont les informations sont présentées. La diversité des indicateurs proposés répond à des besoins variés, mais elle peut aussi rendre la navigation et l'interprétation des données plus complexes pour un utilisateur non familier avec ce type d'outil. En effet, les attentes et les usages des chercheurs diffèrent considérablement selon leur discipline, leur approche méthodologique et leurs besoins analytiques. Une piste d'amélioration à explorer serait de rendre ce tableau de bord personnalisable, en permettant à l'utilisateur de moduler l'affichage des indicateurs selon ses propres besoins. Une telle approche favoriserait une appropriation plus efficace du dispositif et offrirait une expérience plus fluide et adaptée à chaque chercheur. Encore faut-il que cette personnalisation reste accessible : il ne s'agit pas d'attendre de chaque utilisateur qu'il devienne expert de l'outil, mais bien que l'interface propose des modalités d'adaptation simples et intuitives. Une solution intermédiaire pourrait consister à proposer des configurations prédéfinies correspondant à différents profils d'usage, tout en laissant la possibilité à certains utilisateurs, que ce soit individuellement ou pour leur

collectif, d'affiner les paramètres en fonction de leurs besoins spécifiques. Cette logique d'ajustement progressif permettrait d'allier souplesse et lisibilité, sans imposer une charge technique supplémentaire à l'utilisateur final.

Figure 12

Capture d'écran de la page « tableau de bord » du dispositif SoVisu



La liste des dépôts associés au chercheur (figure 13) est organisée en fonction de la métrique de qualité de description (QD) précédemment décrite, qui évalue le niveau de complétude des métadonnées associées à chaque ressource. Cette organisation se décline en trois onglets distincts. Le premier, affiché par défaut, regroupe les dépôts dont les métadonnées sont peu ou pas renseignées. Le second rassemble les dépôts considérés comme bien décrits, c'est-à-dire comportant au minimum un résumé et des mots-clés. Enfin, un troisième onglet offre une vue d'ensemble de l'ensemble des dépôts, sans critère de sélection. L'utilisateur a également la possibilité d'exporter, pour chaque onglet, la liste de ses dépôts sous forme de tableau (CSV) ou BibTeX, facilitant ainsi une intégration dans une bibliothèque Zotero, par exemple.

Figure 13

Capture d'écran de la page « références » du dispositif SoVisu

Références > D. REYMOND

FENÊTRE DE DÉPÔT
1990 - 2025

PROFIL
Vérification des données

SYNTHÈSES
Tableau de bord
Références
Expertise
Lexiques extraits

RESSOURCES
Index des laboratoires
Index des chercheurs

EXPLORATION
Recherche d'expert

PLUS D'INFORMATIONS
Lire la documentation
Présentation
Ressources
Foire aux questions
Déconnexion

Notices à compléter Notices complètes Toutes les notices

Exporter : CSV Zotero Rechercher :

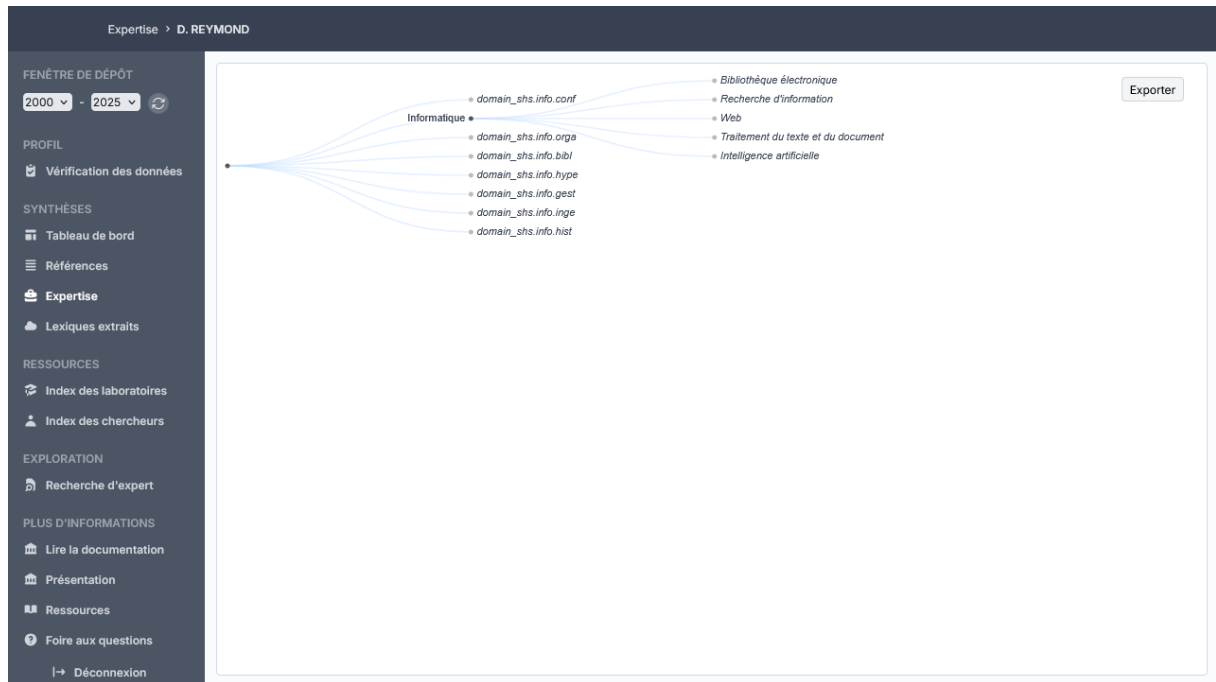
URL	Année	Citations	Titre	QD	Texte complet
hal-01263646	2014	-	Automatic websites classification and retrieval using websites communication signatures	26	
hal-03861913	2022	-	Involving researchers in Open Science. SoVisu: an innovative solution	26	
hal-03975747	2022	-	Projet « relations science industrie : levée d'homonymes inventeurs/auteurs. Patent2net et pivot CIB – bases académiques	26	
hal-00674114	2011	-	Terminologie hypertexte : dynamique temporelle d'une taxinomie.	26	
hal-00716848	2011	-	Représentations informationnelles et dynamiques organisationnelles des sites web : approche temporelle comparative de traces lexicales hypertextuelles	26	
hal-00781585	2014	-	Automatic websites classification and retrieval using websites communication signatures	26	
hal-00674113	2011	-	Une étude terminologique de la communication hypertexte web. Caractéristique du domaine universitaire.	26	
hal-01902242	2016	-	Example of open-source OPS (Open Patent Services) for patent education and information using the computational tool Patent2Net	26	
hal-02064190	2018	-	Quel dommage qu'il y est autant de faute ! La prise en compte des fautes par les consommateurs	26	
hal-02642220	2019	-	Cartographie de l'expertise des chercheurs de l'Université de Toulouse	27	

Affichage de l'élément 1 à 54 sur 54 éléments

La représentation de l'expertise du chercheur (figure 14) est structurée sous forme d'un arbre de compétences à deux niveaux, suivant la classification des domaines d'expertise utilisée sur HAL. Ce modèle hiérarchique permet d'organiser les compétences de manière progressive, des catégories les plus générales aux spécialisations les plus fines. Au premier niveau, on retrouve les grands domaines scientifiques, tels que les sciences humaines et sociales, les sciences de l'ingénieur ou encore la physique. Ces catégories englobent de larges champs disciplinaires et offrent une première classification des expertises du chercheur. Le deuxième niveau affine cette classification en détaillant des sous-domaines spécifiques. Par exemple, au sein des sciences humaines et sociales, on distingue des domaines comme la sociologie, les sciences de l'information et de la communication ou encore l'histoire. Cette segmentation permet d'obtenir une vision plus précise des champs d'étude couverts. Cependant, la principale limite de cette classification à deux niveaux réside dans son manque de granularité. En l'absence d'un troisième niveau, il devient difficile de représenter avec précision certaines expertises pointues ou transversales. Cette structure ne permet pas toujours de refléter pleinement la diversité des compétences des chercheurs, notamment dans des domaines interdisciplinaires ou émergents où une catégorisation plus fine serait nécessaire.

Figure 14

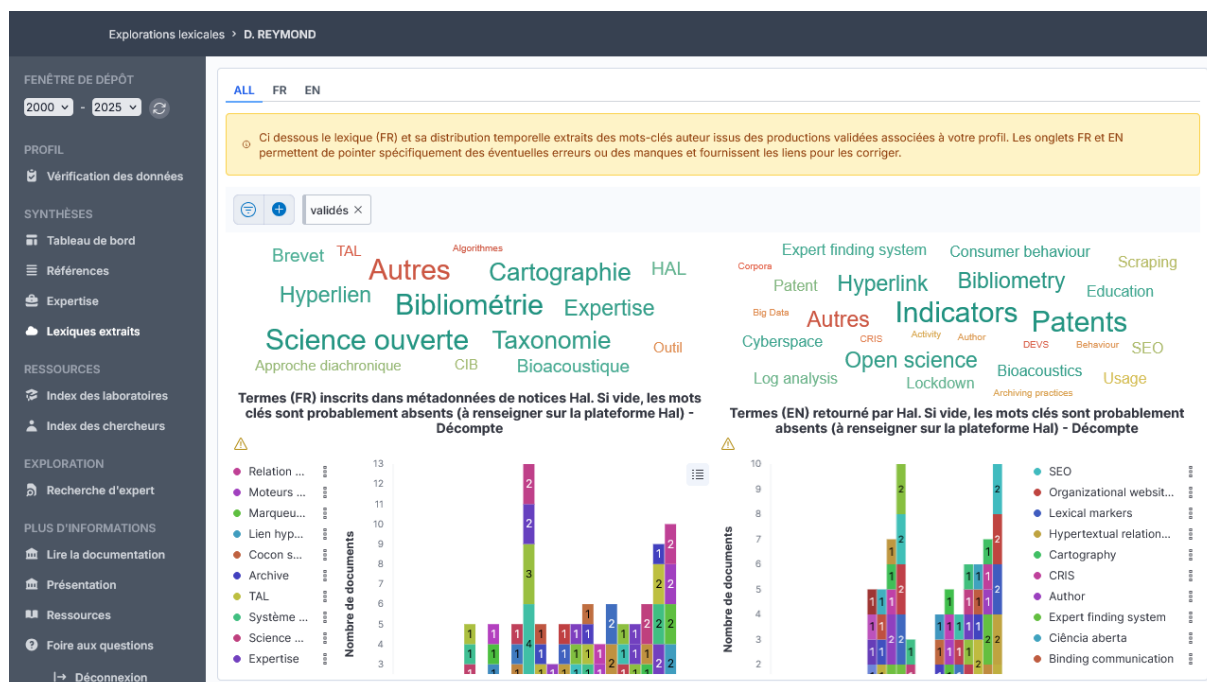
Capture d'écran de la page « expertise » du dispositif SoVisu



En réponse à la limite précédemment formulée, un tableau de bord dédié à l'exploration du lexique extrait des dépôts associés au chercheur (figure 15) permet d'élargir et de préciser la notion d'expertise. En s'appuyant sur une analyse des métadonnées, il offre à l'utilisateur plusieurs visualisations interactives qui révèlent comment son expertise est perçue et interprétée par un système algorithmique, c'est donc sa représentation lexicale qui est présentée. Deux nuages de mots mettent en évidence les termes les plus fréquemment extraits des champs mots-clés et résumés des dépôts de l'auteur. Ces visualisations sont proposées en français et en anglais, permettant d'analyser la terminologie employée dans chaque langue et d'identifier d'éventuelles variations lexicales entre les publications. En complément, deux histogrammes empilés, également déclinés en français et en anglais, illustrent l'évolution temporelle de l'utilisation de ces mots-clés. Cette représentation permet d'identifier les tendances dans les thématiques de recherche abordées par le chercheur au fil du temps, de repérer l'émergence de nouvelles orientations et d'évaluer la stabilité ou l'évolution de son champ d'expertise. Ce tableau de bord propose donc une approche dynamique et contextuelle, offrant une lecture plus fine et nuancée des compétences du chercheur à travers l'analyse des termes qu'il mobilise dans ses travaux.

Figure 15

Capture d'écran de la page « lexique extrait » du dispositif SoVisu



Pour un laboratoire, les liens et finalités présentés dans cette section restent globalement similaires, à l'exception des tableaux de bord, qui sont adaptés pour mieux répondre aux besoins d'un collectif de chercheurs. De plus, un lien « outil » est mis à disposition, offrant des fonctionnalités spécifiques pour faciliter la gestion du laboratoire dans certaines échéances clés.

Les tableaux de bords offrant une visualisation au niveau du laboratoire se déclinent en trois niveaux. Le premier niveau propose une vision d'ensemble des chercheurs rattachés au laboratoire et utilisant le dispositif. Il comprend des tableaux listant les chercheurs avec des informations associées telles que leur numéro d'équipe et le degré de validation de leurs données. Des visualisations complémentaires permettent de représenter ces informations de manière graphique afin d'améliorer la lisibilité. Le second niveau fournit une analyse synthétique des pratiques de dépôt en accès ouvert des chercheurs. Il inclut un diagramme circulaire indiquant le nombre de dépôts identifiés par un DOI, le nombre de documents déposés sur HAL et le nombre de notices sans DOI, qui ne permettent donc pas d'identifier leur statut d'accès libre. Un tableau répertorie les principaux éditeurs auprès desquels les chercheurs publient, tandis qu'un histogramme temporel illustre l'évolution des dépôts affiliés au laboratoire au fil du temps.

Le premier niveau permet d'avoir une vue d'ensemble des chercheurs affiliés au laboratoire qui utilisent le dispositif. Des tableaux listent les chercheurs avec des informations connexes comme le numéro d'équipe et le degré de validation de ses informations. D'autres visualisations viennent décliner sous un aspect visuel ces informations pour des raisons d'accessibilité. Le second niveau offre une visualisation synthétique des pratiques en accès ouvert des chercheurs. Nous retrouvons un diagramme circulaire indiquant le nombre de dépôts identifiés par un DOI, le nombre de documents déposés sur HAL, le nombre de notices sans DOI et par extension, ne permettant pas l'identification de leur statut d'accès libre ou non, un tableau listant les éditeurs principaux dans lesquelles les chercheurs publient et enfin, un histogramme temporel permettant de visualiser l'évolution des dépôts affiliés au laboratoire dans le temps. Le dernier niveau, plus succinct, présente un histogramme temporel permettant de visualiser l'évolution des sources de financement indiquées dans les dépôts affiliés au laboratoire.

Le lien « outil » propose plusieurs fonctionnalités. La première permet d'exporter les données relatives aux laboratoires et aux chercheurs qui y sont affiliés, conformément aux consignes du Hcéres dans le cadre de ses évaluations. Plus précisément, les quatre premiers volets du document demandé par l'autorité sont pré-remplis (figure 16). Cette fonctionnalité représente un gain de temps considérable pour les membres et responsables en charge de la collecte, de la validation et de l'agrégation des données. En générant automatiquement un fichier à partir d'informations déjà validées par les chercheurs, elle assure une fiabilité renforcée et une cohérence accrue dans la production des rapports d'évaluation. La seconde fonctionnalité contribue à renforcer l'intégrité informationnelle en évaluant la cohérence des données entre les profils des chercheurs affiliés et les dépôts associés aux laboratoires. Cette approche permet d'assurer une meilleure interopérabilité des informations dans la base de données.

Figure 16

Capture d'écran d'un résultat d'export Hcéres effectué depuis le dispositif SoVisu

A	B	C	D	E	F	G	H	I	J
authFullName_s	title_s	journalTitle_s	volFull_s	page_s	year	doiId_s	team	Can	utl
VELMURADOVA M.	L'évolution des approches paradigmatiques et théor	Revue française en sciences de l'information et de la communica			2016			N	N
BONJOUR Audrey, MARTIN Fann	Activités de médiation et pratiques langagières dans	SCE-LSA			2016			N	N
BUTTAFOGHI Nicolas, COURBET	Communiquer avec des écrans tactiles. Quelles influ	Revue des Interactions Humaines Médiatisées (RIHM)		25-41	2017		3	N	N
AMIEL Pauline	Le journalisme de solutions, symptôme des mutatio	Questions de communication		307-32	2017	10.4000/q	4	N	N
BONJOUR Audrey, DARAGON Eli	L'assistant digital social et le socionaute	La Revue française de service social			2017		2	N	N
BAAZIZ Abdelkader, LEVEILLÉ Va	Intelligence économique et Développement Durable	Revue Internationale d'Intelligence Économique		57-73	2017		1 ; 2	N	N
BONJOUR Audrey	L'eduction via la médiation technologique	Jeunes et médias -Les cahiers francophones de l'éducation aux			2017		2	N	N
PASCUAL ESPUNY Céline, BONJO	Comment engager pour faire monde commun ? Le	Recherches en communication			2017	10.14428/r	2	N	N
BONJOUR Audrey, DARAGON Eli	Information, formation et mutualisation : Trois défis	Revue Les cahiers de l'Actif			2017		2	N	N
BAIDINO PUTZKA Martine	De Rabelais à Twitter...	Epistémé : revue internationale de scier		3-28	2017			N	N
BERNARD Françoise, DURAMPAR	Créer un co-laboratoire en SIC dans un ancrage régi	Revue française des sciences de l'inform			2017	10.4000/r	2 ; 3	N	N
GOLLIOT Julie, ABELLARD Alexan	Évolution de la Rééducation du Syndrome Dysexé	Transition digitale & médiations numér			2017		3 ; 3	N	N
GERINI Christian, VERDIER Norbe	Rubrique « Textes en questions » : Les textes empru	Quadrature			2017			N	N
HEISER Laurent	L'enaction, un cadre épistémologique fécond et dur	Distances et Médiations des Savoirs			2018			N	N
ZAIONS Ana Paula de Marco Res	Análise da participação brasileira no depósito de pat	Revista Cubana de Información en Cien			2018	10.36512/r	1	N	N
NANNIPIERI Olivier	Agentivité et présence dans les environnements virt	Interfaces numériques			2018	10.25965/r	3	N	N
GALLI David	L'embarras en communication	Hermès, La Revue - Cognition, commun		21-29	2018			N	N
AMATO Stéphane	Décodage émotionnel : arnaque ou cauchemar	Cerveau & Psycho		68-73	2018		2	N	N
FERRAZ Renato Ribeiro Nogueira	Aspectos históricos da criação dos grupos de pesqui	Ciência & Saúde Coletiva		837-84	2018	10.1590/1	1	N	N
BAAZIZ Abdelkader	Smart Muséums : Vers l'émergence de nouvelles m	AIDaInformazion: Rivista di Scienze dell'Informazione		9-36	2018	10.4399/9	1	N	N
AMATO Stéphane, BERNARD Fra	Adhésion, observance et compliance : apports info	MEI - Médiation et information		23-34	2018		2 ; 2	N	N
GALLI David	François Jost, La Méchanceté en actes à l'ère numér	Questions de communication		p. 410	2018	10.4000/questionsde		N	N
GERINI Christian	Matteo Treleani. Qui'est-ce que le patrimoine numér	Questions de communication		472-47	2018	10.4000/questionsde		N	N

Ressources

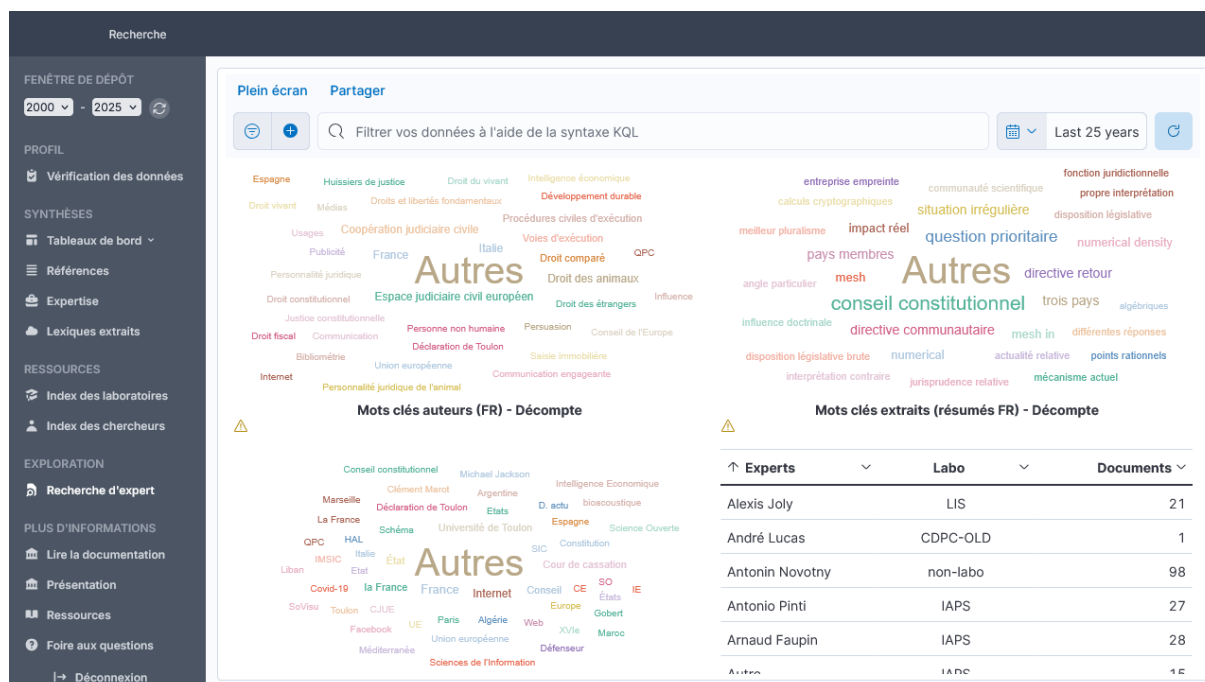
La section « ressources » répertorie deux registres, chercheurs et laboratoires. Ces registres permettent à l'utilisateur d'accéder à des éléments de synthèse, décrits précédemment, à partir de la production scientifique associée à ses collègues, son laboratoire ou d'autres unités de production scientifique. Ils offrent ainsi une vision d'ensemble structurée des contributions scientifiques, facilitant l'exploration, la mise en réseau et l'analyse des dynamiques de recherche.

Exploration

La section « exploration » propose un outil de recherche d'experts (figure 17), permettant à l'utilisateur de saisir un ou plusieurs mots-clés afin d'identifier les chercheurs et laboratoires spécialisés dans un domaine donné. Le tableau de bord affiche alors plusieurs éléments : l'ensemble des mots-clés extraits du champ dédié et des résumés des dépôts correspondant à la requête, les entités nommées détectées dans ces mêmes champs, ainsi qu'un tableau listant les experts identifiés. Ce dernier précise leur affiliation ainsi que le nombre de documents associés à leur nom, offrant ainsi une indication quant au poids de leur expertise sur le sujet recherché.

Figure 17

Capture d'écran de la page « recherche d'expert » du dispositif SoVisu



Plus d'informations

La section « plus d'informations » regroupe plusieurs rubriques permettant d'anticiper les questionnements des utilisateurs et de les accompagner vers de meilleures pratiques en science ouverte. Nous retrouvons la documentation du dispositif SoVisu qui aborde les détails techniques présentés ici. Nous retrouvons également une présentation plus succincte et vulgarisée de l'outil avec un descriptif des étapes clés dans la création du profil chercheur. Nous proposons également des ressources internes à l'Université de Toulon (charte de signature, etc.) et externes pour la gestion des identifiants scientifiques, de la documentation différents aspects afférents à l'utilisation de l'archive HAL, des ressources spécifiques à la science ouverte, notamment les déclarations mentionnées dans le premier chapitre, ainsi que des éléments juridiques liés au dépôt et à la diffusion des travaux scientifiques. Enfin, la foire aux questions répond aux interrogations les plus fréquentes en apportant des éclairages sur les obligations liées au dépôt des travaux, les bénéfices de la science ouverte, la diffusion des publications, ainsi que sur la finalité du dispositif, qui vise avant tout à amener vers une meilleure visibilité de la production scientifique sans être un instrument d'évaluation bibliométrique.

En lien avec ce dernier point, nous avons exploré la possibilité d'intégrer une surcouche interactive qui vient se superposer au dispositif existant. Activable via une combinaison spécifique

de touches, elle permet d'afficher un complément d'information in situ, sans altérer la structure ni l'ergonomie du système initial. Ce type de dispositif se retrouve dans certaines applications dont la compréhension peut s'avérer complexe, comme c'est le cas, par exemple, du Géoportail de l'Institut national de l'information géographique et forestière. Cette approche enrichit l'expérience utilisateur en offrant un accès immédiat et contextuel aux informations additionnelles. En affichant les données in situ uniquement à la demande, elle évite toute surcharge visuelle superflue tout en préservant la logique d'utilisation du dispositif, garantissant ainsi une expérience fluide. Cette surcouche facilite également l'appropriation des contenus en permettant une interprétation directe des informations sans nécessiter de navigation supplémentaire. Pour des problématiques de temps, nous n'avons pas poursuivi l'intégration de cette fonctionnalité.

2.4. Limites

Le dispositif n'a pas uniquement suscité des réactions positives lors de nos communications. Certains chercheurs expriment encore des craintes liées à l'évaluation ou à une forme de « surveillance » induite par les métriques affichées, bien que celles-ci reposent sur des informations déjà accessibles publiquement et couramment utilisées à l'échelle des laboratoires, et non des individus, notamment dans le cadre de l'attribution des bonus de performance mentionnés précédemment. Ce type de réaction n'est pas isolé : il s'inscrit dans un contexte plus large de défiance envers les instruments numériques perçus comme des agents de rationalisation des pratiques scientifiques. D'après Desrosières (2008), toute opération de quantification implique un cadrage politique, un choix de critères, et donc une forme de pouvoir sur ce qui est montré, mesuré, valorisé. En ce sens, SoVisu, même conçu comme outil d'aide, peut être perçu comme un agent de contrôle s'il n'est pas accompagné d'un cadre interprétatif clair et d'une stratégie d'appropriation co-construite avec les usagers. Ce constat renvoie à la nature même de SoVisu, que nous qualifions ici de dispositif au sens fort du terme, c'est-à-dire comme un agencement sociotechnique structurant des relations entre des acteurs, des outils, des normes, des usages et des discours. Dans la lignée des travaux menés au sein du laboratoire IMSIC, un dispositif ne se réduit ni à un outil technique, ni à une procédure : il articule des éléments matériels et symboliques, prescriptifs et interprétatifs, et engage des logiques d'action, de pouvoir et de médiation. Ces réticences peuvent être éclairées par les travaux de Perriault (2008), selon qui un dispositif technique ne vaut que par l'usage qui en est fait ou dévié. La signification d'un outil n'est jamais donnée d'avance, elle se construit dans l'interprétation qu'en fait l'utilisateur. En ce sens, si SoVisu est perçu comme un outil de contrôle, c'est moins en raison de ses fonctionnalités

intrinsèques qu'en raison du contexte d'injonctions multiples, en l'occurrence, performance, visibilité et évaluation, dans lequel le dispositif s'inscrit.

Sur le plan technique, par manque de ressources, nous avons dû faire le choix de ne pas implémenter de solution permettant de suivre et d'analyser finement les interactions des utilisateurs avec le dispositif. Cette limite a restreint notre capacité à documenter les usages effectifs, à identifier les éventuels freins et à ajuster SoVisu en conséquence pour mieux répondre aux attentes des chercheurs. Une démarche centrée sur l'expérience utilisateur, même minimale, aurait néanmoins pu être envisagée en complément, par exemple à travers des entretiens, des observations ou des tests utilisateurs ciblés. Cela aurait permis de recueillir des retours qualitatifs sur les usages réels du dispositif et d'ouvrir la voie à une conception plus participative.

Dans une réflexion plus large, l'intégration d'un nouveau dispositif dans le flux de travail, déjà dense, des chercheurs constitue un défi de taille. L'adoption de nouvelles pratiques, quelles qu'elles soient, nécessite du temps, un accompagnement soutenu et une volonté politique affirmée pour s'ancrer véritablement dans les pratiques. Notre expérimentation avec le dispositif SoVisu a suscité un certain intérêt, mais, pour les raisons évoquées précédemment, n'a pas su convaincre durablement la majorité des chercheurs. Il nous semble pertinent d'établir un parallèle, bien que s'inscrivant dans la thématique plus large de l'archivage, avec les réflexions de Thierry Chanier dans son ouvrage sur les archives ouvertes (Chanier, 2004). Plus de vingt ans après, nombre des obstacles qu'il identifie à leur adoption demeurent encore, malgré les nombreuses initiatives mises en place pour inciter à leur utilisation.

Toutefois, plusieurs années après l'expérimentation, une modification comportementale est identifiable. Ces résultats sont présentés dans le chapitre qui suit, dans la section consacrée à la cartographie des pratiques d'archivage des chercheurs de l'Université de Toulon.

2.5. Conclusion

Ce chapitre explicite en détails les développements successifs du dispositif SoVisu conçu initialement pour accompagner vers de meilleures pratiques dans l'archivage de la production scientifique des chercheurs au sein de l'Université de Toulon. Initialement focalisé sur la visualisation des données moissonnées depuis HAL, le prototype initial s'est enrichi pour devenir un dispositif de médiation documentaire numérique, répondant tant aux besoins stratégiques institutionnels qu'aux attentes spécifiques des chercheurs. Cette approche vise à améliorer la compréhension et la maîtrise par les chercheurs de la découvrabilité de leurs travaux, générant

ainsi un environnement informationnel favorable à leur engagement durable dans la démarche de science ouverte. Dans cette perspective, notre approche s'inscrit pleinement dans le cadre des travaux de Le Deuff (2011), qui insiste sur l'importance d'une solide culture informationnelle comme condition essentielle à une appropriation efficace et durable des outils numériques par les chercheurs. En structurant clairement les informations et en facilitant leur appropriation par les utilisateurs, le dispositif présenté ici contribue activement à la construction de cette culture informationnelle au sein de la communauté universitaire. Cependant, pour garantir une adoption effective à long terme, il est indispensable d'approfondir ces résultats par une analyse qualitative des usages réels et des freins éventuels. Une telle démarche pourrait permettre d'affiner continuellement les fonctionnalités proposées, d'ajuster les stratégies d'accompagnement et ainsi d'optimiser la pertinence du dispositif tant pour les chercheurs que pour les institutions. Sur le plan technique, les pistes d'améliorations sont nombreuses et en partie explorées dans les futures moutures du dispositif présentées sous la dénomination SoVisu+ (Reymond & Lapôtre, 2023). La fonctionnalité la plus importante, selon nous, est d'assurer une intégration bidirectionnelle avec l'archive HAL, afin que toute modification des métadonnées effectuée depuis le dispositif se répercute automatiquement dans l'archive, ce qui n'est pas encore le cas dans le prototype présenté ici. Un tel niveau d'interaction permet un gain de temps notable à l'utilisateur, ce qui constitue une partie de la réponse à une autre problématique majeure dans le cycle de la gestion de l'information scientifique par le chercheur. L'infrastructure technique déployée permet, par ailleurs, d'exposer les données structurées à des interrogations extérieures, offrant ainsi un potentiel intéressant pour proposer un système de recherche d'experts ouvert au public. Cette fonctionnalité favorise de nouvelles formes de collaboration entre chercheurs et acteurs du monde socio-économique, industriels, journalistes, décideurs politiques, en contribuant à diffuser l'expertise scientifique au-delà du strict cadre universitaire.

C'est également à travers cette expérimentation concrète, menée au plus près des pratiques effectives, que plusieurs interrogations centrales de cette thèse ont progressivement émergé. En ce sens, le développement du dispositif SoVisu ne constitue pas uniquement un objet d'étude, mais a contribué à structurer le questionnement scientifique en révélant, sur le terrain, les tensions, besoins et limites liés aux pratiques d'auto-archivage. Ce chapitre, centré sur le dispositif SoVisu, joue ainsi un rôle charnière dans la structure d'ensemble de la thèse. Il articule le contexte institutionnel et politique du libre accès présenté en première partie à une observation empirique des pratiques réelles d'archivage développée dans la troisième. En combinant réflexivité,

développement technique et médiation informationnelle, SoVisu incarne à la fois une réponse concrète aux enjeux de découvrabilité et un levier heuristique pour en renouveler l'analyse. Cette double dimension, à la fois opératoire et analytique, permet d'inscrire pleinement ce travail dans une posture de recherche-action, tout en posant les bases méthodologiques de la cartographie exploratoire menée par la suite.

Après avoir détaillé le processus de conception du prototype et dispositif SoVisu, il apparaît donc essentiel d'explorer plus précisément les pratiques effectives d'archivage sur l'archive ouverte HAL. En effet, si le développement technique et les choix méthodologiques du dispositif visent à faciliter et promouvoir des comportements vertueux d'auto-archivage, leur pertinence ne peut être pleinement évaluée qu'à l'aune des usages réels observés au sein des différentes communautés disciplinaires.

Partie 3 - Cartographie des pratiques d'archivage sur HAL

« L'auto-archivage, loin d'être un automatisme, constitue un choix situé, à la croisée de normes institutionnelles et de préférences individuelles. »

Au cours de ce troisième chapitre, nous explorons notre projet de cartographie des pratiques d'archivage sur l'archive ouverte HAL sous plusieurs perspectives disciplinaires. Ce choix de titre est un dernier clin d'œil à la dénomination initiale de notre premier projet de recherche, CartoHAL, initié il y a quelques années. Dans cette partie, nous adoptons une posture analytique qui fait écho aux travaux de Dominique Cardon (2011), qui invite à penser les données numériques à partir d'un jeu d'échelles : il s'agit de zoomer pour observer les comportements individuels ou localisés, et de dézoomer pour faire émerger des structures collectives et des tendances systémiques. Si le chapitre précédent mobilise principalement une focale resserrée, l'approche développée ici repose davantage sur ce mouvement de « dézoom ». En exploitant les métadonnées à une échelle plus agrégée, nous visons à faire apparaître des dynamiques disciplinaires, voire interdisciplinaires, susceptibles de révéler des logiques globales d'appropriation de l'archive ouverte HAL. Cette circulation entre échelles d'analyse constitue un levier heuristique fort pour comprendre les usages informationnels, tout en rendant visibles les structures qui sous-tendent les pratiques individuelles. Inspirés par Mahé et Prime-Claverie (2017a, 2017b), nous prolongeons ici la réflexion amorcée autour du développement du dispositif d'accompagnement à l'auto-archivage présenté dans le chapitre précédent. Ils offrent une analyse des logiques multiples qui sous-tendent les pratiques d'archivage, afin de dégager des axes d'évolution pertinents pour ce dispositif. Nous présentons d'abord la méthodologie employée, avant de vous inviter à explorer notre cartographie des pratiques d'archivage. Cette exploration débute par les sciences, techniques et médecine, se poursuit avec les spécificités des sciences humaines et sociales, pour enfin mettre l'accent sur notre discipline, les sciences de l'information et de la communication. Enfin, nous analysons le cas de l'Université de Toulon en le comparant à celui d'universités de taille similaire, afin de mettre en évidence les inférences du déploiement du dispositif SoVisu sur les pratiques d'archivage propres à l'établissement toulonnais.

3.1. Méthodologie

3.1.1. L'analyse quantitative

Contrairement aux sciences, techniques et médecine, où l'approche du quatrième paradigme (Hey et al., 2009) est largement admise pour l'étude des phénomènes à travers l'analyse massive des données, l'usage des méthodes quantitatives demeure moins répandu en sciences humaines et sociales. Pourtant, avec l'émergence des sciences sociales de troisième génération (Boullier, 2017), fondées sur l'exploitation des « traces numériques », un nouveau méthodologique se dessine. Ces traces, loin d'être de simples reflets du réel, participent activement à la structuration des normes et des pratiques scientifiques (Latour & Biegunski, 2005).

Dans ce cadre, l'analyse des interactions en ligne permet d'explorer des phénomènes complexes à une échelle inédite, dépassant les limites des approches traditionnelles. Les sciences de l'information et de la communication, en lien étroit avec les technologies de l'information et de la communication, sont particulièrement concernées par ces évolutions. L'étude des dynamiques d'archivage dans des plateformes comme HAL illustre cet intérêt : les dépôts d'articles et les interactions des chercheurs avec la plateforme, au travers des métadonnées associées constituent des marqueurs informationnels pertinents pour comprendre les pratiques de diffusion scientifique. Notre approche s'inscrit dans l'évolution des paradigmes méthodologiques des sciences de l'information et de la communication, qui tendent à dépasser l'analyse instrumentale de la recherche d'information pour apprécier une vision plus large des pratiques informationnelles. Comme le soulignent Chaudiron et Ihadjadene (2010), ces pratiques doivent être envisagées dans leur dimension sociale, située et contextualisée. En ce sens, l'analyse quantitative que nous menons sur les dépôts HAL ne se limite pas à une lecture statistique des actes de dépôt ou de complétude des métadonnées : elle vise à interroger les logiques d'usage, les routines et les arbitrages des acteurs. Cette posture rejoint celle défendue par Paquien-séguy (2012), qui insiste sur la nécessité de considérer les usages numériques à travers les cadres techniques, les dispositifs médiateurs et les structures institutionnelles qui les conditionnent.

Toutefois, l'adoption d'une approche quantitative ne signifie pas l'abandon des méthodes qualitatives, bien au contraire. Didier Courbet (Courbet, 2013) souligne l'importance du pluralisme méthodologique dans la discipline, qui permet non seulement d'élargir le spectre des objets de recherche mais aussi de renforcer la validité des connaissances produites. Loin d'une opposition rigide entre quantitatif et qualitatif, cette complémentarité méthodologique favorise une approche plus systémique des phénomènes étudiés.

Ainsi, nous adoptons une approche quantitative, centrée principalement sur l'analyse des statistiques descriptives, afin d'étudier les contributions sur l'archive ouverte HAL. Nous considérons ces contributions comme des traces numériques, laissées par les chercheurs, bibliothécaires et autres acteurs de la recherche, permettant d'examiner les dynamiques d'archivage et les pratiques associées. Cette approche rejoint les réflexions de Merzeau (2013) sur la traçabilité comme construction socio-technique, où les métadonnées de HAL révèlent moins des comportements individuels que des négociations collectives entre normes disciplinaires et injonctions institutionnelles. Lorsque cela s'avère heuristique, et dans la logique de complémentarité méthodologique mentionnée précédemment, nous articulons nos analyses avec des études qualitatives qui enrichissent l'interprétation des résultats. Cette approche nous permet

non seulement d'identifier des tendances d'ensemble, mais aussi de contextualiser ces dynamiques à la lumière des pratiques et des stratégies des acteurs impliqués.

3.1.2. *Objectif de recherche*

L'objectif de cette recherche est de cartographier et d'analyser les pratiques d'auto-archivage des chercheurs sur la plateforme HAL, en prenant en compte les différentes logiques de dépôt et les typologies de contributeurs. Cette démarche s'inscrit dans une approche exploratoire inspirée du paradigme méthodologique proposé par Schmitt (2018), où l'exploration et la visualisation de grandes masses de données constituent des leviers centraux pour produire des connaissances nouvelles à partir des traces numériques. En effet, si l'archive ouverte joue un rôle majeur dans la diffusion des résultats scientifiques, elle ne se limite pas à une simple fonction de stockage mais participe à l'organisation et à la structuration du savoir scientifique, comme l'a souligné Jean Meyriat dans ses travaux sur les systèmes documentaires. Les pratiques d'archivage hétérogènes témoignent ainsi des enjeux liés à la formalisation des connaissances et aux modes de circulation de l'information scientifique. Dans cette perspective, cette étude accorde une attention particulière aux spécificités disciplinaires qui façonnent les usages de l'auto-archivage. Comme le rappellent Chartron et Schöpfel (2017), chaque discipline développe des pratiques distinctes en matière d'accès ouvert, à l'intersection de ses traditions communicationnelles, de ses enjeux scientifiques et de ses préoccupations professionnelles. L'analyse des comportements de dépôt sur HAL permet ainsi de mieux saisir les dynamiques propres à chaque communauté de recherche. Ce questionnement est d'autant plus pertinent que, dans un nombre croissant d'établissements et de structures de recherche, le dépôt dans HAL tend à devenir une exigence institutionnelle, inscrite dans les politiques de valorisation scientifique ou les procédures d'évaluation. Cette étude se propose donc de répondre à plusieurs questions clés : qui dépose sur l'archive HAL et comment ? Quels moyens peuvent être mis en œuvre pour améliorer ces pratiques de dépôt tout en garantissant une organisation cohérente et structurée de l'information ? Et dans l'observation d'une prescription locale : quelle est l'efficacité d'un dispositif d'accompagnement à la science ouverte tel que celui développé au sein de l'Université de Toulon ?

Nous approfondirons ainsi l'analyse des quatre logiques de dépôt identifiées par Mahé et Prime-Claverie (communication scientifique directe, référencement, archivage, recensement), en examinant leur impact sur la visibilité des travaux. Nous chercherons également à caractériser les profils des contributeurs au travers de leurs pratiques d'archivage. Cette recherche contribue ainsi à une meilleure compréhension des dynamiques sous-jacentes aux contributions dans les archives

ouvertes tout en éclairant les voies possibles pour améliorer les stratégies de diffusion scientifique.

3.1.3. *Périmètre de recherche*

En France, l'archive ouverte HAL joue un rôle central dans la diffusion de l'information scientifique. Confirmée par les politiques nationales en faveur de la science ouverte, la plateforme permet aux chercheurs, bibliothécaires et autres acteurs de la recherche d'archiver leurs travaux, quelle que soit la discipline dont ils relèvent. L'étude de ces pratiques de dépôt, à l'origine même de l'acte d'archiver, constitue un élément essentiel pour comprendre l'adoption et l'évolution du libre accès dans l'hexagone. En ce sens, pour cartographier ces pratiques et analyser leurs dynamiques associées, notre périmètre de recherche s'étend à l'intégralité des notices déposées sur la plateforme du premier janvier 2002 au 31 décembre 2023. Nous pourrions ainsi en référer à différentes périodes des politiques ministérielles en faveur de la science ouverte, en particulier celles qui ont marqué l'évolution des pratiques d'archivage.

Nous prenons le choix d'exclure les notices issues d'imports automatisés. Ce choix vise à concentrer notre analyse sur les pratiques d'archivage manuelles et effectuées directement sur la plateforme HAL.

Les pratiques d'archivage diffèrent, entre autres, en fonction des politiques universitaires et des disciplines d'appartenance des auteurs. Nous analysons donc ici plusieurs sous-ensembles de notices déposées, à savoir, un ensemble de notices relevant des sciences, techniques et médecine, un relevant des sciences humaines et sociales, un autre relevant des sciences de l'information et de la communication et, enfin, un ensemble de notices associées à l'Université de Toulon. Ces corpus sont décrits plus amplement par la suite.

Nous étudions le premier sous-ensemble avec l'hypothèse que ce segment présente une adoption précoce et plus systématique des pratiques de science ouverte et par extension d'archivage, un trait qui semble courant dans les disciplines techniques et médicales où les exigences des journaux internationaux et des organismes de financement encouragent une large diffusion des résultats de recherche. L'adoption précoce de ces pratiques peut être en partie attribuée à l'émergence d'arXiv qui a joué un rôle important dans la promotion de l'archivage dans bon nombre de disciplines relevant des sciences, techniques et médecine.

En contraste, le second sous-ensemble peut montrer des variations significatives en raison de la grande dispersion éditoriale (Chanier, 2004) et de la moindre influence des sociétés savantes, ce

qui rend la standardisation des pratiques de dépôt plus difficile. En sciences humaines et sociales, chaque publication tend à être un cas particulier en raison de l'hétérogénéité des formats et des exigences éditoriales : « chaque revue représente presque un cas particulier », écrit Chanier, rendant ainsi difficile l'établissement de modèles types et, par conséquent, une moindre propension au libre accès comparée aux disciplines scientifiques et techniques.

Le troisième sous-ensemble, qui concerne les dépôts en sciences de l'information et communication, est présumé manifester des pratiques exemplaires en termes de découvrabilité et de gestion des métadonnées. En plus d'intégrer les sciences de l'information et des bibliothèques (*Library and Information Science*) depuis Jean Meyriat (Meyriat & Miège, 2002), les chercheurs de la discipline travaillent au cœur des nouvelles technologies de l'information. Cette expertise devrait se refléter dans leurs pratiques de publication et d'archivage.

Le sous-ensemble de notices associées à l'Université de Toulon offre une perspective sur l'influence des politiques universitaires locales sur les pratiques d'archivage au sein d'une institution pluridisciplinaire. L'étude de ce dernier sous-ensemble permet particulièrement d'évaluer le degré d'efficacité du dispositif SoVisu évoqué précédemment sur ces pratiques.

À des fins de comparaison, des corpus similaires sont établis selon les critères et méthodes décrites ci-après, à partir des notices produites du premier janvier 2002 au 31 décembre 2022. Bien qu'utilisés de manière limitée dans les analyses présentées ici, ces derniers sont documentés et mis à disposition pour des recherches futures (Tabariès, 2024a).

Cette démarche vise notamment à inviter à étendre l'étude avec une éventuelle analyse de l'historicité des modifications des dépôts entre les deux périodes moissonnées, en accord avec l'idée de suivre l'évolution des documents et des pratiques de dépôt au fil du temps. Cette approche permettrait d'étudier la manière dont les chercheurs mettent à jour leurs dépôts et comment les corrections, ajouts, ou retraits de documents sont effectués.

Il convient de souligner que les pratiques d'archivage observées dans cette partie sont étroitement liées aux évolutions techniques, réglementaires et organisationnelles de la plateforme HAL elle-même. À cela s'ajoutent des obstacles d'ordre organisationnel, identifiés dans l'enquête de Schöpfel et al. (2022), qui pointent notamment le manque d'accompagnement structuré et la faible valorisation institutionnelle du dépôt dans certaines unités de recherche. En ce sens, certains changements de comportement repérés chez les déposants peuvent être lus en miroir des

grandes étapes du développement de HAL, comme l'introduction du dépôt automatique par les éditeurs, la mise en place de portails institutionnels ou l'évolution des formulaires de dépôt.

3.1.4. Caractérisation des dépôts et déposants

Nous exploitons les données de la plateforme HAL en interrogeant l'interface de programmation mise à disposition, puis, en calculant des champs, dans un programme exécutable écrit dans le langage de programmation Python. Les données moissonnées comportent une partie des métadonnées associées et, de ce fait, qui décrivent les notices déposées sur l'archives.

Nous effectuons des traitements basiques pour faciliter l'analyse des données. Nous vérifions la présence de champs descriptifs (fichiers, mots-clés, résumé) indépendamment du langage et nous attribuons une métrique mesurant le niveau de description de la ressource en question (Tabariès, 2022). Cette approche s'inscrit dans les réflexions de Jean Meyriat (Meyriat, 1981) sur l'importance des systèmes documentaires comme vecteurs de structuration du savoir scientifique. En effet, la richesse et la précision des métadonnées ne se limitent pas à un rôle purement technique ; elles déterminent la manière dont l'information scientifique est organisée, retrouvée et mobilisée au sein des communautés de recherche. Notre approche de la qualité des métadonnées s'inscrit également dans le cadre théorique développé par Simonnot (2012) sur l'accès à l'information en ligne. Selon cette perspective, la qualité des métadonnées ne constitue pas seulement un enjeu technique mais représente un élément structurant dans la visibilité et l'accessibilité des productions scientifiques. Par ailleurs, comme le souligne Mayère (2010), ces pratiques de description s'inscrivent dans des logiques organisationnelles qui influencent la manière dont les chercheurs documentent leurs travaux, révélant ainsi les tensions entre les exigences institutionnelles et les pratiques effectives.

Pour des raisons de praticité, nous étudions uniquement la présence du champ testé, et non la qualité même de l'information qu'il contient. Ainsi, notre « score de découvrabilité » ne mesure pas l'exactitude des métadonnées, mais leur complétude formelle, en tant que condition de leur repérabilité et de leur réutilisation. Pour affiner cette analyse, nous appliquons un filtre spécifique aux résumés : lorsque tous les résumés, quelle que soit la langue, sont plus courts que le titre de la notice, ou bien contiennent des expressions standardisées telles que « no abstract », la ressource est considérée comme dépourvue de résumé. Ce score repose sur un calcul pondéré à partir de plusieurs champs jugés essentiels à la visibilité de la ressource (titre, résumé, mots-clés, DOI, texte intégral, etc.). Un premier calibrage, publié sous la forme d'un « data paper », dans la Revue française des sciences de l'information et de la communication (Tabariès, 2022), proposait une

pondération différenciée selon l'importance estimée de chaque champ. Cette approche, bien que heuristique, posait des problèmes d'arbitrage interprétatif. Nous privilégions désormais une pondération uniforme, plus lisible et plus équitable, pour permettre une comparaison cohérente entre les notices.

Chaque notice est enrichie par les métriques d'usage et d'impact internes à HAL, c'est-à-dire les nombres de visionnage et de téléchargement, en moissonnant directement la page web de la notice. Lorsqu'un identifiant DOI est renseigné pour une notice, nous interrogeons l'interface de programmation fournie par Dimensions pour récupérer le nombre de citations et le rapport relatif des citations associés (*field citation ratio*), c'est-à-dire le nombre de citations qu'une publication reçoit par rapport à la moyenne des citations reçues par d'autres publications similaires dans le même domaine scientifique et publiées la même année. Bien que le moissonnage des *altmetrics* à l'aide de l'interface de programmation proposée par le site éponyme et leur étude soit indéniablement intéressante, très peu de travaux référencés sur l'archive en disposent. Nous avons donc choisi de ne pas les inclure dans notre méthode.

Nous nous sommes basés sur les travaux de Mahé et Prime-Claverie (2017a, 2017b) pour calculer les champs décrivant les niveaux de contribution, les logiques de dépôt et les types de contributeurs pour chaque dépôt, en adoptant une méthodologie entièrement automatisée du fait de la taille des corpus.

Les niveaux de contribution

Les niveaux de contributions décrivent la personne à l'origine même de l'acte de dépôt. Nous retrouvons trois niveaux possibles, du moins éloigné au plus éloigné de l'objectif initial de communication scientifique directe :

- le premier niveau, **auto-archivage**, ou « **communication scientifique directe entre chercheurs** », lorsqu'un dépôt est effectué par un auteur du document, on parle d'auto-archivage. Le calcul de ce champ s'appuie sur le champ *selfArchiving_bool* fourni par HAL pour chaque notice. Nous vérifions si la chaîne de caractère inscrite dans le champ contributeur correspond au nom d'un des auteurs, et nous corrigeons la valeur dans le cas où l'information fournie par HAL s'avère erronée ;
- le second niveau, « **intermédiaire au service de la communauté scientifique** », lorsqu'un dépôt est effectué par un intermédiaire, c'est-à-dire, par une personne qui n'est pas inscrite comme auteur du document. Le calcul de ce champ se base sur l'étude du nom du contributeur. En s'appuyant sur la librairie *names-dataset* (2022), nous comparons

la chaîne de caractère inscrite dans le champ contributeur à un ensemble de noms et de prénoms issus d'une fuite de données du réseau social numérique Facebook. Si la chaîne de caractères s'apparente au nom d'une personne physique, nous considérons que le dépôt est effectué par un intermédiaire, le cas échéant, nous considérons que le dépôt est effectué dans le cadre d'un transfert massif de données (voir ci-après) ;

- le dernier niveau, « **intermédiaires extérieurs au service des institutions** », lorsqu'un dépôt est effectué dans le cadre d'un transfert massif de notices à partir d'une autre plateforme.

Nous reprenons donc partiellement la typologie telle que définie par Mahé et Prime-Claverie. Pour des contraintes techniques, nous abstrayons la différenciation entre les contributions effectuées par des intermédiaires chercheurs de celles effectuées par des intermédiaires non chercheurs. Nous pensons que l'introduction d'erreurs liées à des problèmes d'homonymie sur l'étude d'un ensemble de documents aussi conséquent et les possibles changements dans le statut professionnel des contributeurs, lesquels pourraient avoir effectué des dépôts en leur qualité de chercheurs par le passé ne permettent pas d'analyser ce critère de manière rigoureuse. Il est également important de noter que les chercheurs ne représentent pas toujours les principaux contributeurs aux dépôts dans HAL. Comme l'indiquent Schöpfel et al. (2023) une part importante des dépôts, jusqu'à 48 % dans certains cas, est réalisée par des bibliothécaires ou du personnel administratif, illustrant un déplacement du modèle d'auto-archivage initial vers une gestion institutionnelle plus structurée

Les logiques de dépôt

Les logiques de dépôt permettent de différencier les dépôts récents et anciens et avec ou sans texte intégral associé et traduisent la logique selon laquelle le dépôt est effectué. Nous reprenons la typologie et les méthodes de calcul définies par Mahé et Prime-Claverie, nous retrouvons donc quatre logiques de dépôt.

À savoir qu'une notice suit la première logique de « **communication scientifique directe** » lorsqu'elle est créée au plus tard l'année suivant la publication du document et qu'un document est associé à cette dernière. Lorsque la notice n'est pas associée à un document, elle suit la seconde logique de « **visibilité de front de recherche (référencement)** ».

Une notice suit la troisième logique d'« **archivage** » lorsqu'elle est créée au moins une année après la publication du document et qu'un document est associé à cette dernière. Lorsque la

notice n'est pas associée à un document, elle suit la quatrième logique de « **visibilité des recherches antérieures (recensement)** ».

Les logiques de communication scientifique directe et d'archivage, de part la présence d'un document associé, participent directement à la diffusion de l'information scientifique à l'inverse des deux autres logiques. Il est également important de noter que les pratiques de dépôt dans HAL sont marquées par des disparités disciplinaires et institutionnelles. Ces différentes logiques de dépôt interrogent la dimension réflexive du geste d'auto-archivage évoquée précédemment. Lorsqu'aucun fichier n'est associé à la notice, il ne s'agit pas nécessairement d'un oubli, mais d'un positionnement : entre désengagement, prudence ou stratégie d'exposition. En ce sens, l'auto-archivage, loin d'être un automatisme, constitue un choix situé, à la croisée de normes institutionnelles et de préférences individuelles. Mahé et Prime-Claverie (2017a) montrent que, dans le domaine des sciences humaines et sociales, les chercheurs dépendent souvent de politiques institutionnelles pour s'engager dans l'auto-archivage. Ces politiques influencent la fréquence et le type de dépôts effectués, soulignant l'importance des incitations locales

Les types de contributeurs

Les types de contributeurs décrivent la fréquence et la continuité des dépôts effectués par les contributeurs sur l'archive. Ils définissent les habitudes de contribution. Toujours en s'inspirant des travaux de Mahé et Prime-Claverie, nous définissons quatre catégories différentes :

- la première catégorie représente les **nouveaux contributeurs pour plusieurs années**. Ils n'ont jamais déposé auparavant mais déposent plusieurs travaux, au moins, dans les deux années suivantes le dépôt originel ;
- la seconde catégorie représente les **contributeurs réguliers**. Ils déposent au moins une fois tous les deux ans ;
- la troisième catégorie représente les **contributeurs occasionnels** qui déposent une première fois mais ne contribuent pas dans les fenêtres observées deux ans avant et après le dépôt originel ;
- la quatrième catégorie représente les **contributeurs sortants** qui ont déposé auparavant mais qui n'ont pas effectué de nouveaux dépôts au cours des deux années suivantes.

Ces catégorisations diffèrent légèrement des travaux initiaux de Mahé et Prime-Claverie. Contrairement aux auteurs qui observent les comportements sur une fenêtre de une année, nous observons ces comportements sur une fenêtre de deux ans. Nous choisissons d'augmenter la fenêtre d'observation pour mieux saisir les cycles de publication plus longs, plus adaptés aux

sciences humaines et sociales. Nous ne caractérisons pas les dépôts des années N-1 et N-2 en raison de la durée de la période d'observation définie.

Au final, pour chaque notice, nous ajoutons les champs présentés dans le tableau 1. Les métadonnées moissonnées depuis HAL et encore présentes dans les jeux de données sont listées dans le tableau 2.

Tableau 1

Métadonnées ajoutées aux notices HAL

Intitulé	Type	Description
<i>harvested_on</i>	Date	Date de récolte de la notice (ISO 8601)
<i>metrics_harvested_on</i>	Date	Date de récolte des métriques (ISO 8601)
<i>has_file</i>	Booléen	Indique la présence d'un fichier
<i>has_abstract</i>	Booléen	Indique la présence d'un résumé
<i>has_keywords</i>	Booléen	Indique la présence de mots-clés
<i>contributor_type</i>	Texte	Niveau de contribution
<i>deposit_logic</i>	Texte	Logique de dépôt
<i>behavior</i>	Texte	Type de contributeur
<i>field_citation_ratio</i>	Décimal	Rapport relatif des citations
<i>times_cited</i>	Entier	Nombre de citations
<i>times_viewed</i>	Entier	Nombre de visionnages (métrique interne à HAL)
<i>times_downloaded</i>	Entier	Nombre de téléchargements (métrique interne à HAL)
<i>qd</i>	Décimal	Mesure de « découvrabilité » (niveau de description ; [0, 1])

Tableau 2

Métadonnées moissonnées depuis HAL

Intitulé	Type	Description
<i>submittedDate_tdate</i>	Date	Date de dépôt (ISO 8601)
<i>publicationDate_tdate</i>	Date	Date de publication électronique (ISO 8601)
<i>selfArchiving_bool</i>	Booléen	Dépôt en auto-archivage
<i>contributorFullName_s</i>	Texte	Nom complet du contributeur/déposant
<i>authFullName_s</i>	Texte	Auteur : Nom complet
<i>domain_s</i>	Texte	Codes domaines du document
<i>fileMain_s</i>	Texte	URL du fichier principal

3.1.5. Infrastructure et aspects techniques

Pour appliquer la méthode décrite précédemment, nous avons déployé une infrastructure capable de gérer et d'analyser un volume important de données issues de l'archive étudiée. Cette infrastructure repose sur des conteneurs Docker, une technologie permettant un déploiement cohérent et reproductible des applications, indépendamment de l'environnement sous-jacent. Dans un souci de transparence et de reproductibilité, tous les scripts développés sont déposés sur HAL (Tabariès, 2024b), en complément de l'infrastructure détaillée ci-après et illustrée dans la figure 18.

La première étape consiste à interroger l'interface de programmation de recherche HAL (*API HAL | API Archive Ouverte HAL*, s. d.). La requête suivante, utilisée pour interroger cette interface est présentée ci-après.

```
https://api.archives-ouvertes.fr/search/?q=*&fl=$flags&start=$increment&rows=$rows&fq=submittedDate_tdate:[$gte TO $lt]&sort=docid%20asc
```

Cette requête est exécutée de manière incrémentale en utilisant les paramètres *start* et *fq*, afin de moissonner l'ensemble des notices disponibles au moment de la requête initiale. Ce procédé assure une couverture exhaustive de l'archive. L'association des paramètres *fq* et *submittedDate_tdate* permet une segmentation des requêtes par années, facilitant ainsi le contrôle du processus. Cette approche garantit également la possibilité de reprendre le moissonnage à un point précis en cas d'interruption ou de problème technique. Après le moissonnage des données,

nous effectuons divers calculs pour caractériser les dépôts, en suivant les méthodes détaillées précédemment. Ces données sont ensuite téléversées vers une instance Elasticsearch (*elastic/elasticsearch*, 2010/2024), qui sert de base de données et de moteur d'indexation. Elasticsearch s'impose comme une solution idéale pour notre projet en raison de sa capacité à traiter de grands volumes de données, de sa rapidité d'indexation et de recherche, ainsi que de sa scalabilité. Ces atouts sont essentiels pour analyser les plus de trois millions de notices étudiées. D'autant plus que, récemment, la suite Elasticsearch a retrouvé son statut de logiciel en code source ouvert (Banon, Shay, 2024). Un processus d'automatisation complémentaire fonctionne en tâche de fond afin de récupérer les métriques d'impact et d'analyser la présence ou l'absence d'autres dépôts sur une période donnée. L'exploration et l'analyse des données moissonnées s'effectuent à l'aide de Kibana (*elastic/kibana*, 2013/2024), une extension à Elasticsearch offrant une interface interactive et intuitive. Grâce à ses fonctionnalités de visualisation dynamique, Kibana facilite l'identification des tendances au sein de vastes corpus de données. Une illustration des capacités de Kibana est disponible dans la figure 19, où un tableau de bord montre un exemple d'exploration des données.

Figure 18

Processus et infrastructure de collecte

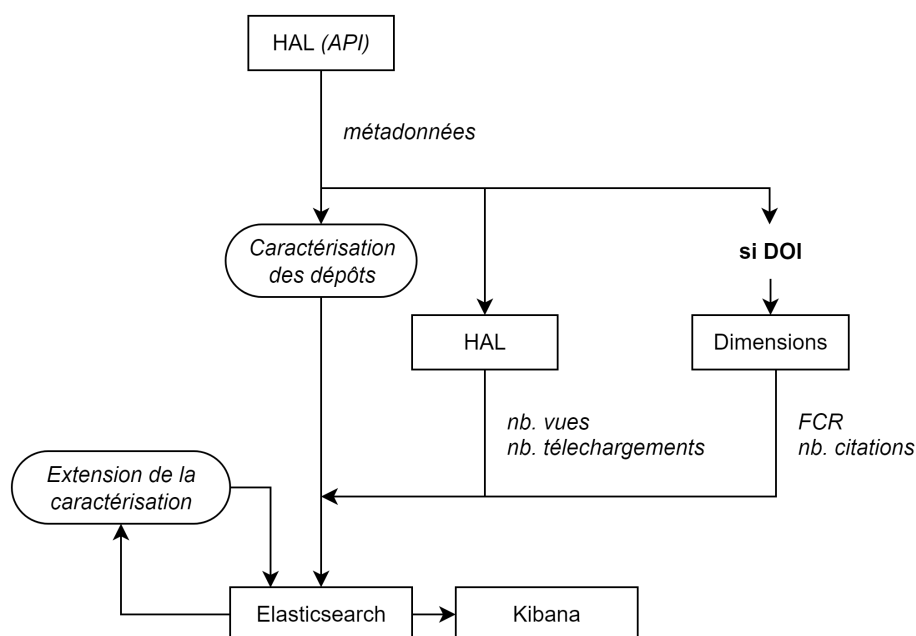
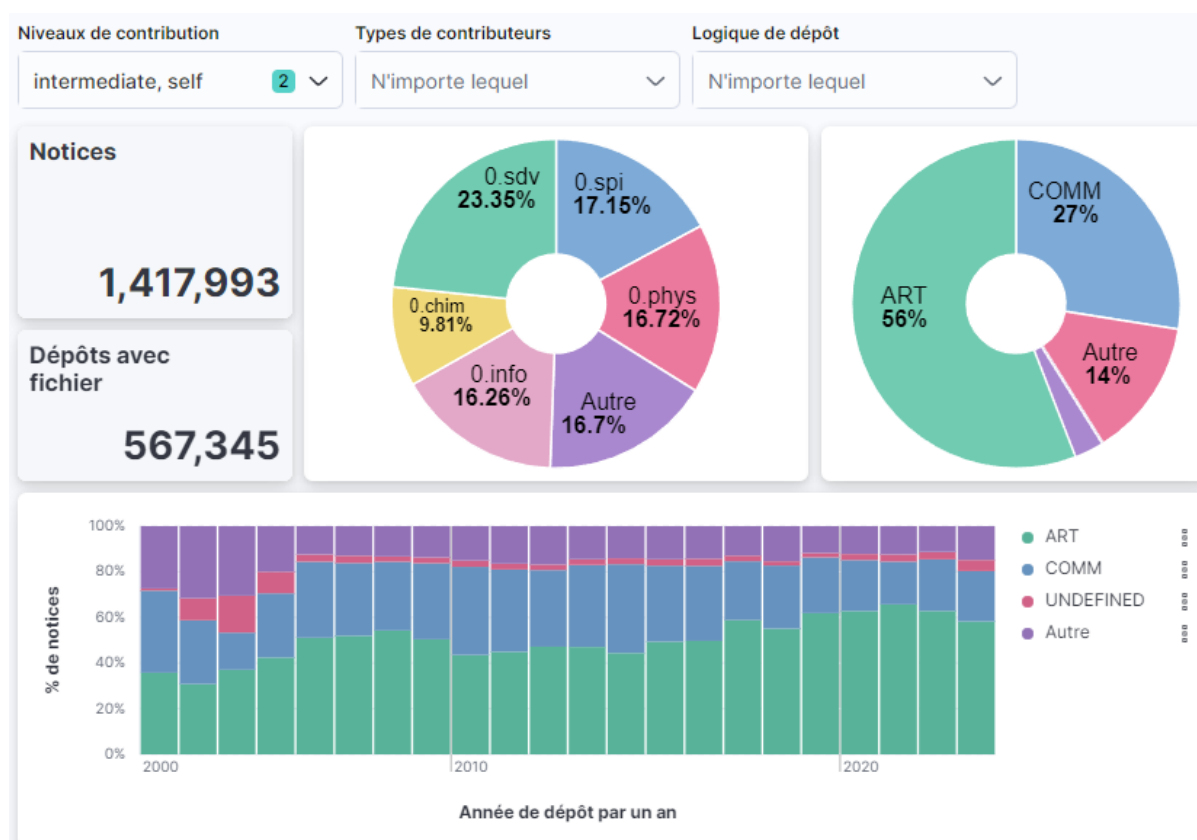


Figure 19

Capture d'écran d'une partie d'un tableau de bord Kibana



3.1.6. Description des corpus

Les différents corpus évoqués précédemment, à savoir, un corpus comprenant les notices HAL en sciences, techniques et médecine, un autre en sciences humaines et sociales, ainsi que celui en sciences de l'information et de la communication, et enfin un corpus de notices tamponnées par l'Université de Toulon offrent un panorama diversifié de la production scientifique archivée sur la plateforme. Ces ensembles présentent des caractéristiques élémentaires différentes. Nous proposons ci-après une présentation de ces informations pour chaque instantané, capturés à la fin de l'année 2022, puis à la fin de l'année 2023. Pour rappel, les notices issues d'imports automatisés ne sont pas comprises dans les corpus, ce qui explique les volumes de notices inférieurs aux nombres mis en avant sur les différents portails.

Corpus de notices HAL en sciences, techniques et médecine

Pour l'année 2022, le corpus de notices HAL en sciences, techniques et médecine se compose de 1 297 405 notices, dont 499 913 sont accompagnées de fichiers (38,5 %). Un total de 79 404 contributeurs différents ont participé à ces dépôts, soit une moyenne de 16,35 notices par

contributeur. Les répartitions disciplinaires révèlent une diversité relative, avec 22,4 % des dépôts dans les sciences de la vie, 17,31 % en sciences de l'ingénieur, 17,21 % en physique, 16,84 % en informatique, et 9,75 % en chimie. Les articles représentent 56 % des dépôts, suivis des communications à 28 %.

En 2023, ce corpus s'étend à 1 417 993 notices, dont 567 345 présentent un fichier lié (40 %). Le nombre de contributeurs est passé à 87 735, soit une moyenne de 16,17 notices par contributeur. Les sciences de la vie dominent toujours avec 23,35 % des dépôts, suivies des sciences de l'ingénieur (17,15 %), de la physique (16,72 %), de l'informatique (16,26 %) et de la chimie (9,81 %). Les articles constituent toujours 56 % des dépôts, tandis que les communications représentent 27 % et les autres types de documents 14 %.

Corpus de notices HAL en sciences humaines et sociales

En 2022, le corpus de notices HAL en sciences humaines et sociales comprend 861 196 notices, dont 231 705 avec un fichier lié (26,9 %), et 41 627 contributeurs uniques, soit une moyenne de 20,69 notices par contributeur. Les domaines de l'histoire (14,61 %), du droit (9,18 %), de la littérature (8,71 %), de l'archéologie (8,27 %), et de la sociologie (7,53 %) sont les plus représentés. Les articles constituent 36 % des dépôts, suivis des communications et des chapitres d'ouvrage, chacun à 21 %.

En 2023, ce corpus s'étend à 999 211 notices, avec 261 755 fichiers liés (26,2 %), et 48 388 contributeurs différents, soit une moyenne de 20,65 notices par contributeur. Les dépôts sont principalement en histoire (14,19 %), droit (8,98 %), littérature (8,8 %), archéologie (7,8 %) et sociologie (7,17 %). Les sciences de l'information et de la communication ne représentent que 3,6 % de ce corpus. La proportion des articles reste à 36 %, les communications et chapitres d'ouvrage à 21 % chacun, avec d'autres types de documents à 22 %.

Corpus de notices HAL en sciences de l'information et de la communication

Le corpus de 2022 pour les sciences de l'information et de la communication compte 32 067 notices, dont 12 454 avec fichier lié (38,8 %), déposées par 4 234 contributeurs, soit une moyenne de 7,57 notices par contributeur. Les articles représentent 33 % des dépôts, suivis des communications (30 %) et des chapitres d'ouvrage (15 %).

En 2023, ce corpus s'étend à 36 056 notices, avec 13 452 fichiers liés (37,3 %) et 4 591 contributeurs, soit une moyenne de 7,85 notices par contributeur. Les articles forment 32 % des

dépôts, les communications 30 %, les chapitres d'ouvrage 15 %, et d'autres types de documents 22 %.

Corpus de notices HAL tamponnées Université de Toulon

En 2022, le corpus de notices tamponnées par l'Université de Toulon comprend 23 397 notices, avec 7 560 fichiers liés (32,3 %), déposées par 2 872 contributeurs, soit une moyenne de 8,14 notices par contributeur. Le droit (18,81 %), la physique (13,84 %), l'informatique (13,32 %), les sciences de l'ingénieur (11,59 %) et les sciences de l'information et de la communication (8,13 %) dominant. Au niveau des champs disciplinaires, 56 % des notices relèvent des sciences humaines et sociales alors que 44 % relèvent des sciences, techniques et médecine. Les articles représentent 54 % des dépôts, les communications 25 %, les chapitres d'ouvrage 8 %, et les autres types 13 %.

Pour 2023, ce corpus présente 25 437 notices, avec 9 060 fichiers liés (35,6 %) et 3 159 contributeurs, soit une moyenne de 8,05 notices par contributeur. Les principales disciplines restent le droit (18,57 %), la physique (13,57 %), l'informatique (13,14 %), les sciences de l'ingénieur (11,45 %) et les sciences de l'information et de la communication (7,93 %). La répartition des champs disciplinaires reste inchangée. Les articles représentent 53 % des dépôts, les communications 25 %, les chapitres d'ouvrage 8 %, et les autres types 14 %.

Pour relativiser les résultats présentés ci-après, il nous semble utile de mentionner quelques répartitions disciplinaires des jeux de données utilisés en tant qu'éléments de comparaison sans pour autant y consacrer un paragraphe dédié. À savoir que, le jeu de données propre aux notices tamponnées par l'Université d'Avignon comporte 33 % de notices relevant des sciences, techniques et médecine et 67 % relevant des sciences humaines et sociales. Le corpus intégrant l'intégralité des notices HAL comporte autant de notices relevant des sciences, techniques et médecine que de notices relevant des sciences humaines et sociales.

3.1.7. *Accès aux données*

Les données utilisées dans cette thèse sont publiées et librement accessibles sur la plateforme Nakala (Tabariès, 2024a), toujours dans une démarche de transparence et de reproductibilité. Les données déposées sur l'archive sont exportées depuis une instance Elasticsearch (*elastic/elasticsearch*, 2010/2024) en utilisant l'outil *Elasticdump* (*elasticsearch-dump/elasticsearch-dump*, 2013/2024), la méthode d'importation est documentée sur le répertoire cité. En raison de contraintes juridiques, les données propres aux métriques de citations, récoltées par l'intermédiaire de Dimensions, sont retirées. Le jeu de données contient les schémas des deux

index (2022 et 2023) et les données associées compressées au format GZip. Suivant l'exemple de la licence CC0 appliquée aux métadonnées de HAL, ces jeux de données sont également distribués sous la même licence CC0, permettant leur utilisation, modification et partage sans aucune restriction.

3.2. Résultats

Dans cette section, nous présentons les résultats de notre cartographie des pratiques d'archivage des chercheurs et personnels de la recherche sur l'archive ouverte HAL. Les données moissonnées et analysées couvrent la période de 2002 à fin 2023, offrant une vue d'ensemble des tendances et dynamiques d'archivage. Les résultats présentés ci-après reprennent en partie ceux publiés dans le dossier numéro 24 de la revue française des sciences de l'information et de la communication (Tabariès, 2022) et à la 23ème édition du Colloque International sur le Document Electronique 23 (Tabariès & Reymond, 2023).

Nous rappelons que nos hypothèses de recherche postulent des variations significatives des pratiques d'archivage selon les disciplines, avec une adoption plus systématique et précoce des pratiques de science ouverte dans les domaines des sciences, techniques et médecine. À l'inverse, nous anticipons que les sciences humaines et sociales montrent une diversité de pratiques du fait d'une plus grande dispersion éditoriale et d'exigences moins homogènes. Les sciences de l'information et de la communication, en raison de leurs particularités, sont supposées présenter des différences par rapport à leur domaine parent. Enfin, l'étude des notices tamponnées par l'Université de Toulon nous permet d'évaluer l'efficacité d'un dispositif local d'accompagnement à la science ouverte.

Nous structurons la présentation et l'analyse des résultats ci-après en suivant l'ordre établi lors de la formulation des hypothèses. Une cartographie étant intrinsèquement visuelle et pour fournir une compréhension claire et approfondie des données recueillies, nous avons intégré un nombre important de graphiques dans notre analyse. Ces visuels illustrent les tendances, les comparaisons et les corrélations qui émergent de notre étude.

3.2.1. *Les sciences, techniques et médecine*

Dans le cadre de cette analyse des sciences, techniques et médecine, nous entreprenons une exploration approfondie des dynamiques de dépôt et de contribution. Nous commençons par examiner l'évolution du nombre de dépôts et du nombre de contributeurs pour établir une vue

d'ensemble des tendances de participation et mesurer leur impact sur le corpus des archives. Cette analyse nous permet de tracer le portrait des variations dans le temps et d'identifier les facteurs influents. Dans le prolongement, nous étudions l'évolution des niveaux de contribution, discernant l'auto-archivage des dépôts effectués par des intermédiaires, pour voir comment ces différents niveaux influencent les pratiques et la gestion des archives. Poursuivant notre exploration, nous nous intéressons aux différents types de contributeurs, distinguant les nouveaux, les réguliers, les ponctuels et les sortants. Cette distinction nous aide à comprendre les comportements variés au sein de la communauté scientifique et les motivations qui sous-tendent leur engagement dans le processus de dépôt. Nous abordons ensuite les logiques de dépôt, cherchant à comprendre les stratégies adoptées par les contributeurs et comment celles-ci ont évolué au fil du temps. Cette partie de l'analyse révèle comment les pratiques de gestion et de partage de l'information scientifique changent et influencent la nature et la qualité des dépôts. Nous approfondissons notre étude en examinant les logiques de dépôt selon le niveau de contribution, distinguant l'auto-archivage des dépôts effectués par des intermédiaires. Cette analyse détaille comment l'autonomie des contributeurs affecte leur manière de participer aux archives, soulignant les nuances dans les pratiques selon l'autonomie de chacun. De même, nous étudions les logiques de dépôt selon les types de contributeurs, mettant en lumière les différences entre les entrants et les sortants. Cette approche permet de comprendre l'impact de ces différents groupes sur les stratégies de dépôt et sur la gestion de l'archivage. Ensuite, notre attention se porte sur l'évolution du niveau de description des dépôts, de l'absence de description aux dépôts accompagnés de résumés et de mots-clés. Cette observation est importante pour évaluer la qualité et la trouvabilité de l'information archivée. Nous précisons également comment les niveaux de description varient en fonction des logiques de dépôt, notamment entre le recensement et le référencement, dont les intentions sous-jacentes à l'acte de dépôt influencent le processus même de dépôt. Enfin, nous examinons les variations des niveaux de description selon le niveau de contribution et les types de contributeurs. Cette exploration révèle comment l'engagement des contributeurs et leur rôle au sein de la communauté modulent la richesse des informations métadonnées associées aux dépôts.

Les figures 20 et 21 illustrent respectivement l'évolution du nombre de dépôts et du nombre de contributeurs uniques dans le domaine des sciences, techniques et médecine (STM) de 2002 à 2023.

En observant ces deux graphiques, trois phases distinctes se dégagent, caractérisant l'adoption et l'intensification des pratiques d'archivage au sein de cet ensemble de disciplines scientifiques.

La première phase, de 2002 à 2008, se distingue par une croissance modérée et régulière tant du nombre de dépôts que du nombre de contributeurs. Cette période initiale semble marquée par une sensibilisation progressive aux pratiques d'archivage, avec un nombre limité de contributeurs et de dépôts, indiquant une phase d'adoption encore embryonnaire.

La deuxième phase, de 2009 à 2016, est caractérisée par une croissance modérée et fluctuante du nombre de dépôts et une augmentation progressive mais plus lente du nombre de contributeurs. Cette période pourrait indiquer une phase de consolidation des pratiques d'auto-archivage, où l'élargissement de la base des contributeurs s'accompagne d'un rythme de dépôt plus irrégulier mais soutenu. Cette dynamique reflète une intégration croissante des chercheurs dans les processus d'archivage, malgré une absence de croissance exponentielle à ce stade.

Enfin, la troisième phase, à partir de 2017 jusqu'à 2023, est marquée par une accélération exponentielle tant du nombre de dépôts que du nombre de contributeurs. Ce bond significatif coïncide avec les initiatives gouvernementales en faveur de l'archivage ouvert, et peut également refléter une prise de conscience accrue de l'importance de la diffusion libre des résultats de recherche. Il s'inscrit aussi dans un contexte d'amélioration continue de la plateforme HAL, qui a notamment simplifié son interface de dépôt afin de faciliter l'engagement des chercheurs. Notamment, l'année 2023 atteint un sommet historique, avec plus de 150 000 dépôts et plus de 25 000 contributeurs. Cela illustre une certaine phase de maturité dans l'adoption des pratiques d'archivage dans le domaine STM.

Ces tendances soulignent non seulement l'élargissement de la participation dans la communauté scientifique, mais aussi l'importance croissante accordée à la diffusion du savoir scientifique, dans un contexte où l'archivage ouvert s'affirme comme une voie centrale pour y parvenir.

Figure 20

Évolution de la quantité de dépôts dans le temps en STM

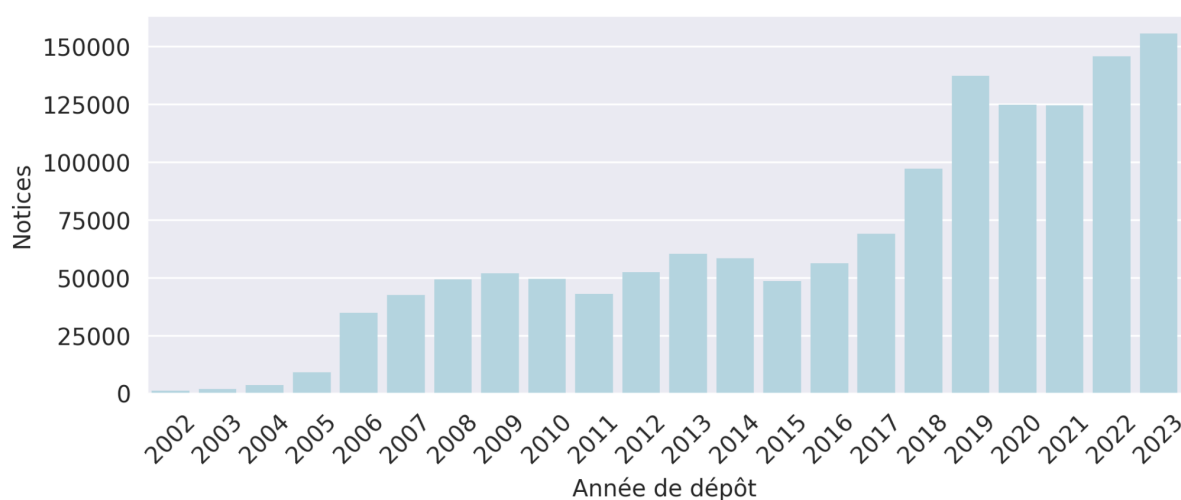
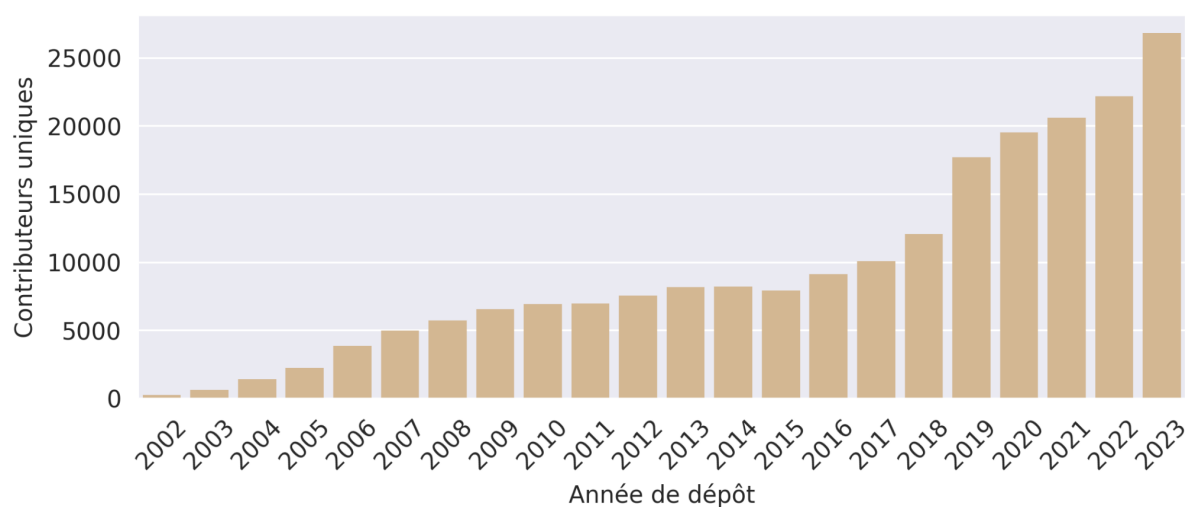


Figure 21

Évolution du nombre de contributeurs uniques dans le temps en STM



La figure 22 présente l'évolution des différents types de contributeurs dans le domaine des sciences, techniques et médecine de 2002 à 2021. Pour rappel, cette analyse distingue quatre types de contributeurs : les réguliers, les occasionnels, les nouveaux, et les sortants.

L'observation des courbes permet de discerner plusieurs dynamiques significatives dans la participation à l'archivage.

Les contributeurs réguliers (ligne bleue) montrent une croissance continue et marquée, atteignant un pic important en 2021 avec près de 10 000 contributeurs. Cette augmentation soutenue

suggère que la base des contributeurs établis, qui participent de manière récurrente à l'archivage, s'est développée au fil des années, reflétant une appropriation durable des pratiques d'auto-archivage.

Les contributeurs occasionnels (ligne verte) suivent une tendance similaire jusqu'à 2018, après quoi leur nombre commence à diminuer légèrement. Cette baisse peut indiquer une transition de certains contributeurs occasionnels vers un statut plus régulier ou une moindre participation ponctuelle dans un contexte où les efforts d'archivage se structurent davantage.

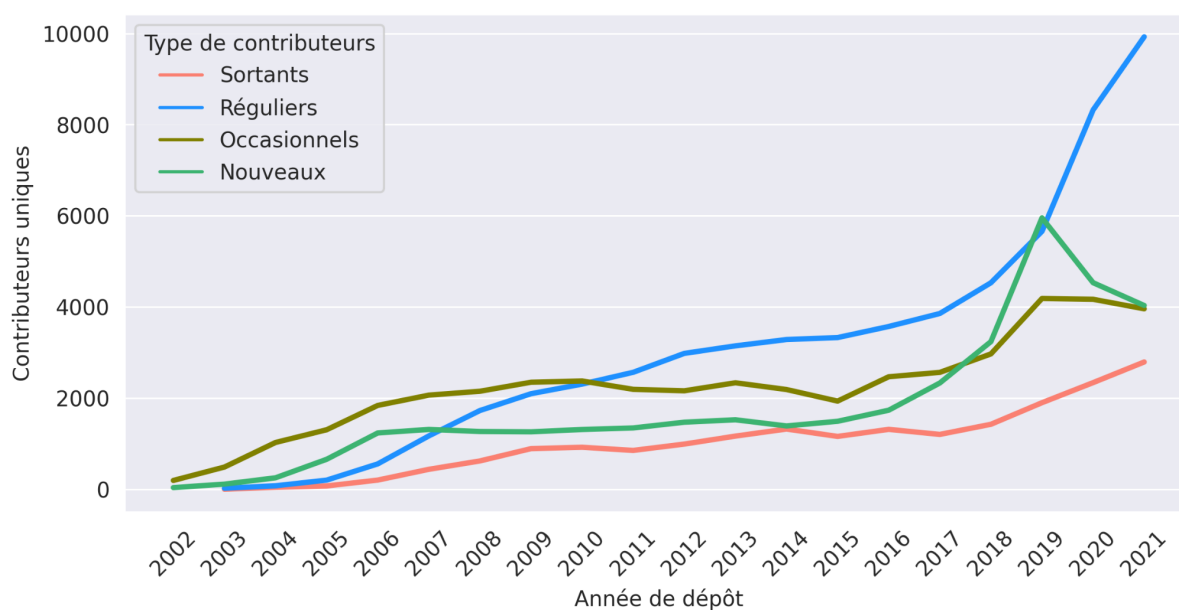
Les nouveaux contributeurs (ligne orange) présentent une courbe de croissance prononcée à partir de 2016, culminant en 2019 avant de connaître une légère diminution. Cela reflète l'intégration continue de nouveaux participants dans le processus d'archivage, bien que le ralentissement récent puisse suggérer une saturation partielle ou une diminution de l'apport de nouveaux contributeurs.

Enfin, les contributeurs sortants (ligne rouge) augmentent également et logiquement progressivement, atteignant un sommet en 2021.

Cette tendance pourrait refléter un renouvellement partiel au sein de la communauté, où une proportion significative de contributeurs cesse de participer activement après une période d'engagement.

Figure 22

Évolution des types de contributeurs dans le temps en STM



La figure 23 illustre l'évolution des logiques de dépôt dans le domaine des sciences, techniques et médecine (STM) entre 2002 et 2023.

À travers cette représentation, nous observons la transformation des pratiques d'archivage au sein de la communauté scientifique. Au début de la période, la communication scientifique directe domine largement, avec un pic notable en 2003 où cette logique représente plus de 60 % des dépôts. Cette forte prévalence témoigne d'une volonté marquée de diffuser rapidement les résultats scientifiques, en associant systématiquement un texte intégral aux dépôts. Cependant, dès 2005, un changement notable s'opère. La part de la communication scientifique directe diminue rapidement, cédant progressivement du terrain aux logiques de recensement et de référencement. Cette période marque une transition vers des pratiques d'archivage plus orientées vers la visibilité et la documentation des recherches antérieures, au détriment d'une diffusion immédiate. À partir de 2011, nous entrons dans une phase de stabilisation des logiques de dépôt.

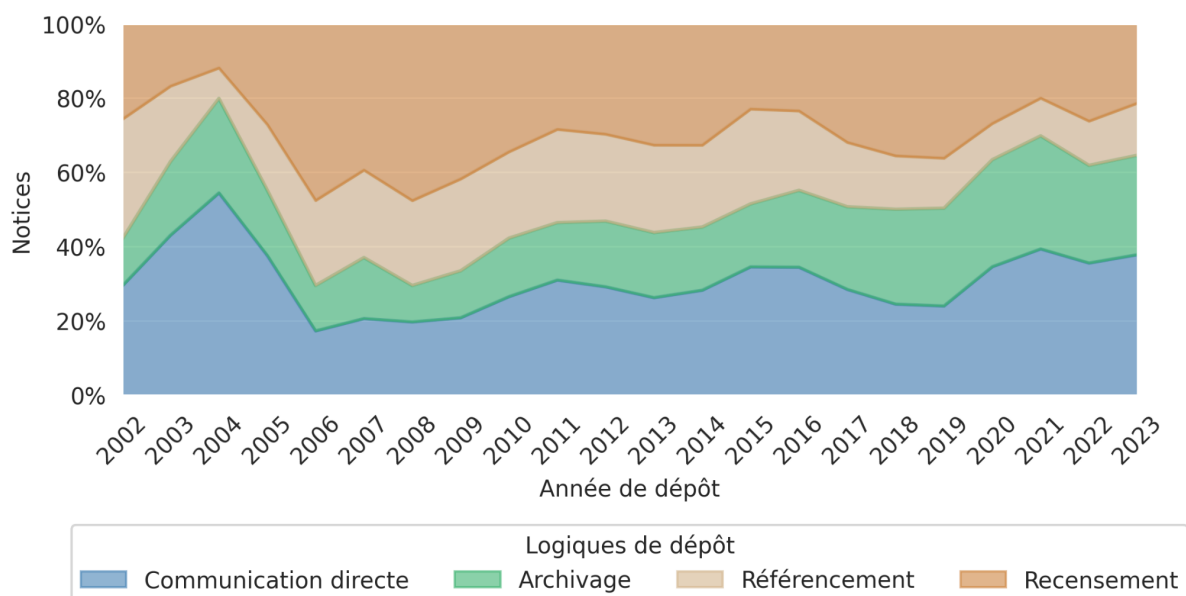
L'archivage, qui consiste à déposer des travaux plus anciens accompagnés d'un texte intégral, gagne en importance au fil des années, jusqu'à représenter environ 26,8 % des dépôts en 2023.

Parallèlement, la communication scientifique directe, après avoir chuté, reprend doucement de l'ampleur, atteignant 37,8 % la même année. Le recensement, qui avait pris une place prédominante dans la décennie précédente, voit sa part se stabiliser tout en restant une des logiques majeures, tandis que le référencement se maintient autour de 14 % des dépôts. Cette évolution des pratiques révèle une diversification et une maturation des stratégies d'archivage au sein des disciplines STM.

Alors qu'au début des années 2000, l'accent était largement mis sur la diffusion immédiate des résultats scientifiques, les années suivantes ont vu émerger une approche plus diversifiée. Celle-ci intègre non seulement la diffusion rapide mais aussi la conservation à long terme et la visibilité des recherches passées. Cette transformation témoigne d'une adaptation progressive des chercheurs aux exigences croissantes de pérennité et de traçabilité des données scientifiques, où chaque logique de dépôt joue désormais un rôle complémentaire dans la gestion et la valorisation des connaissances scientifiques.

Figure 23

Évolution des logiques de dépôt dans le temps en STM



La figure 24 présente l'évolution des logiques de dépôt dans les domaines sciences, techniques et médecine (STM), avec une distinction entre les contributeurs en auto-archivage (haut) et les contributeurs intermédiaires (bas), de 2002 à 2023.

Pour les contributeurs en auto-archivage, la période 2002-2023 se distingue par une variation notable des différentes logiques de dépôt. Initialement, la logique de communication scientifique directe domine largement, représentant plus de 69 % des dépôts en 2002. Cependant, cette logique décline progressivement au profit des autres, atteignant environ 40 % en 2023. La logique d'archivage suit une trajectoire inverse, avec une tendance à la hausse, passant de 30 % en 2002 à environ 25 % en 2023, après avoir atteint un pic autour de 28 % en 2020. La logique de recensement connaît une augmentation plus lente mais continue, culminant à 33 % en 2019, puis redescendant légèrement à 20 % en 2023. Enfin, la logique de référencement, qui reste la logique de dépôt la moins courante, varie peu au fil des ans, représentant environ 13 % des dépôts en 2023, après un pic en 2007 et 2008 autour de 18 %.

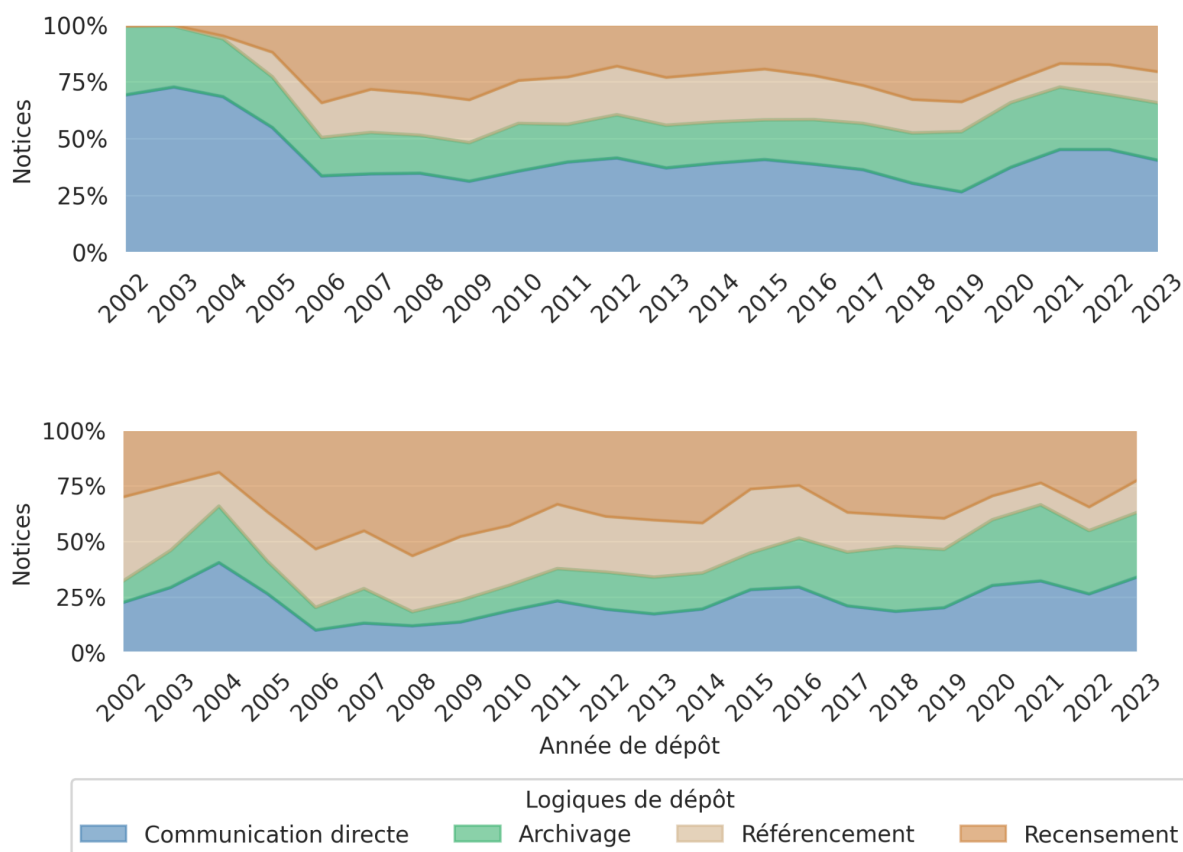
Pour les contributeurs intermédiaires, l'évolution des logiques de dépôt révèle des dynamiques différentes. La logique de recensement occupe une place prédominante tout au long de la période, atteignant un pic à plus de 56 % en 2008, avant de redescendre à 22 % en 2023. La logique de communication scientifique directe présente également une forte présence, débutant à environ 22 % en 2002 et fluctuant avant de culminer à 34 % en 2023. La logique d'archivage, quant à elle,

suit une tendance similaire à celle observée pour les contributeurs en auto-archivage, progressant régulièrement pour atteindre environ 29 % en 2023. La logique de référencement reste stable dans le temps, bien qu'elle occupe une place plus importante en début de période, avec près de 38 % en 2002, avant de chuter à environ 14 % en 2023.

En comparant les deux niveaux de contribution, il apparaît que les contributeurs en auto-archivage tendent à privilégier la logique de communication scientifique directe de manière plus marquée que les contributeurs intermédiaires. Les contributeurs intermédiaires sont, en revanche, plus enclins à suivre une logique de recensement dans le processus de dépôt. Ces différences montrent que les contributeurs intermédiaires privilégient la visibilité des recherches antérieures alors que les contributeurs en auto-archivage adoptent une approche plus diversifiée dominée par une volonté de diffusion rapide des résultats récents. Toutefois, cette tendance semble commencer à s'inverser à la fin de la période observée, les contributeurs intermédiaires semblent tendre progressivement vers plus de dépôts suivant des logiques de communication scientifique directe et d'archivage. Selon nous, ce changement récent pourrait être étroitement lié à l'attribution progressive de rôles de gestionnaires de collections HAL dans les diverses entités de la recherche, en particulier au sein des laboratoires.

Figure 24

Évolution des logiques de dépôt dans le temps en STM déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)



La figure 25 illustre l'évolution des logiques de dépôt dans les domaines sciences, techniques et médecine (STM), avec une distinction entre les contributeurs réguliers (haut) et les nouveaux contributeurs (bas), sur la période allant de 2004 à 2021.

Pour les contributeurs réguliers, la logique de communication scientifique directe domine progressivement, passant de 41 % en 2004 à 43,57 % en 2021. Cela montre une forte préférence pour des dépôts effectués rapidement après publication et accompagnés d'un document, favorisant ainsi une diffusion immédiate de l'information scientifique. Cependant, cette tendance connaît quelques fluctuations, notamment entre 2005 et 2017, où la part de cette logique oscille entre 20 % et 37 %. La logique d'archivage, bien qu'inférieure à celle de communication directe, connaît également une progression régulière au fil des années, passant de 15,19 % en 2004 à 29,69 % en 2021. Ce schéma suggère que les contributeurs réguliers s'efforcent de combler progressivement le retard dans le dépôt de leurs travaux passés, renforçant ainsi la conservation à long terme des publications. La logique de recensement, quant à elle, décroît de manière notable,

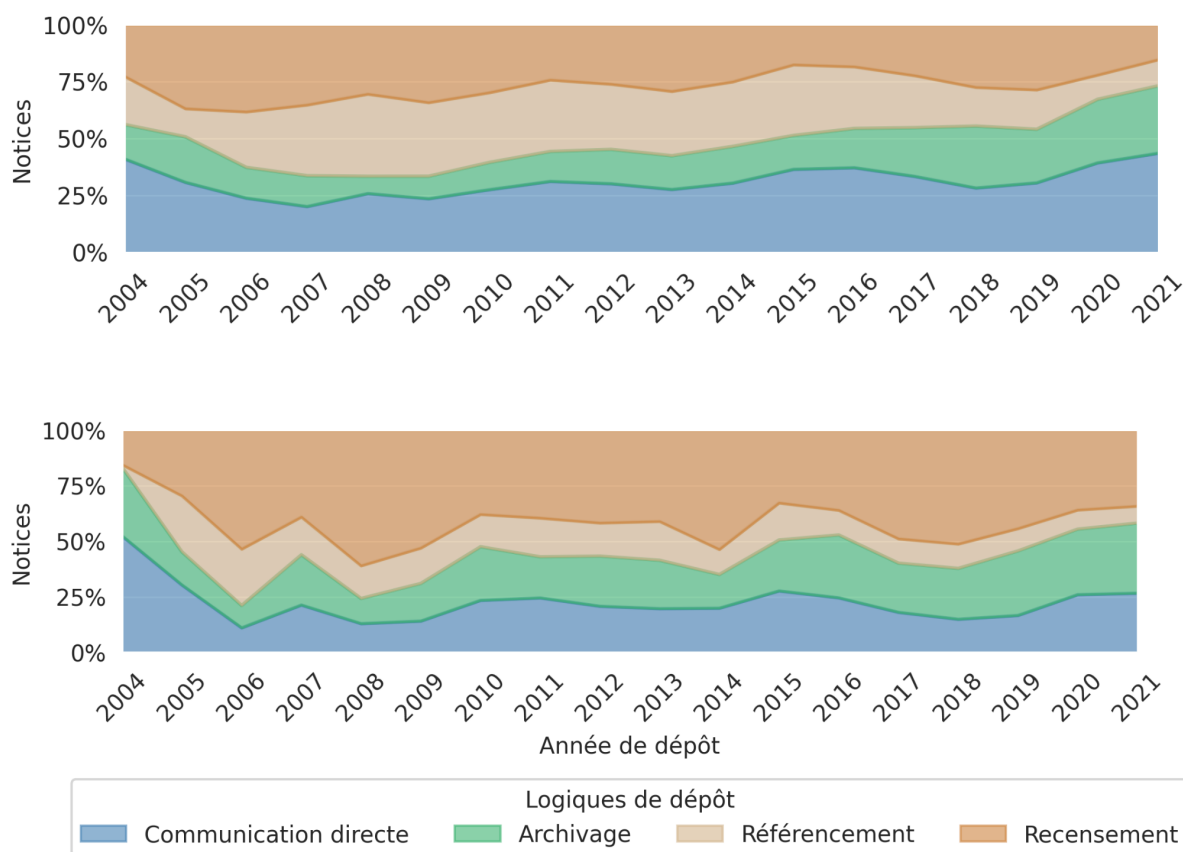
représentant 22,66 % en 2004, puis chutant à 15,33 % en 2021. Cela traduit un intérêt déclinant pour la simple visibilité des travaux sans document associé. Enfin, la logique de référencement varie au cours des années, atteignant des sommets en 2007 (31 %) avant de retomber à 11,39 % en 2021.

Pour les nouveaux contributeurs, la dynamique est légèrement différente. La logique de communication scientifique directe occupe également une place prépondérante, notamment au début de la période avec un pic à 52,26 % en 2004, avant de diminuer régulièrement pour se stabiliser autour de 26,59 % en 2021. Cette baisse peut s'expliquer par une familiarisation progressive avec les pratiques d'archivage et un recours plus fréquent à d'autres logiques de dépôt. La logique de recensement reste une logique dominante chez les nouveaux contributeurs, atteignant 61,09 % en 2008 et restant significatif tout au long de la période. Cependant, cette tendance diminue légèrement à partir de 2015 pour atteindre 34,31 % en 2021. Contrairement aux contributeurs réguliers, les nouveaux contributeurs semblent accorder plus d'importance à la simple visibilité des recherches passées sans document associé. La logique de référencement, bien qu'elle connaisse une certaine importance en début de période, notamment en 2005 (25,09 %), tend à s'effacer progressivement, avec une part de seulement 7,54 % en 2021.

En comparant les deux types de contributeurs, on observe que les contributeurs réguliers tendent à privilégier la communication scientifique directe et l'archivage, des logiques de dépôt qui favorisent la diffusion immédiate et la conservation à long terme des documents. En revanche, les nouveaux contributeurs sont plus enclins à utiliser la logique de recensement, qui n'inclut pas de document associé, ce qui traduit une approche davantage orientée vers la visibilité des travaux sans diffusion immédiate. La logique de référencement est également plus fréquente chez les nouveaux contributeurs en début de période, mais son utilisation diminue au fil des ans dans les deux groupes. Les contributeurs réguliers s'orientent donc plus vers une gestion plus anticipée et complète des publications, tandis que les nouveaux contributeurs semblent adopter des pratiques moins complètes axées sur une simple visibilité.

Figure 25

Évolution des logiques de dépôt dans le temps en STM déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)



La figure 26 retrace l'évolution des niveaux de description des dépôts dans le domaine des sciences, techniques et médecine (STM) de 2002 à 2023, mettant en lumière les changements dans les pratiques d'archivage au fil des ans. Le graphique distingue trois niveaux de description : les dépôts accompagnés de mots-clés et de résumés (en vert), ceux accompagnés de mots-clés ou de résumés (en jaune), et les dépôts sans description (néant, en rouge).

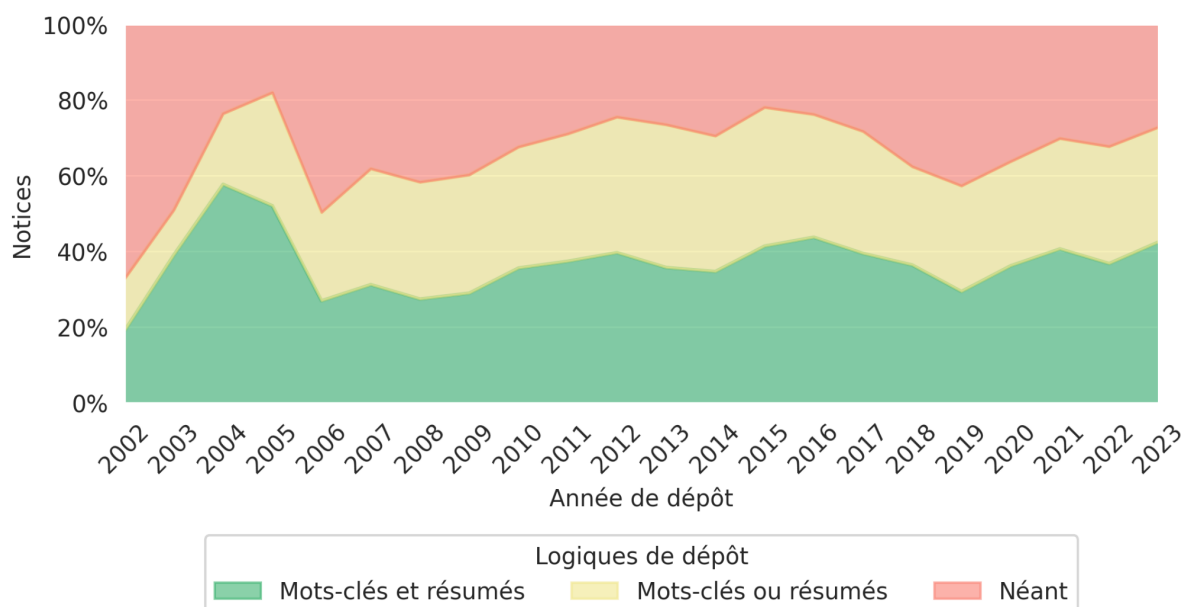
Dès le milieu des années 2000, les pratiques d'archivage commencent à se structurer. Alors qu'une part importante des dépôts était initialement faite sans aucune description, cette tendance commence à s'inverser rapidement à partir de 2004. Les dépôts accompagnés de mots-clés et de résumés voient leur proportion augmenter significativement, représentant environ 58 % des dépôts en 2004. Cependant, entre 2006 et 2016, les pratiques montrent une certaine variabilité. La proportion de dépôts sans description connaît un regain en 2006, atteignant près de 50 %, suggérant que malgré les progrès, les pratiques d'archivage ne sont pas encore entièrement stabilisées. Cette période est marquée par des fluctuations, où l'équilibre entre le renseignement

de métadonnées complètes et la simple présence de mots-clés ou de résumés reste incertain. À partir de 2017, une tendance plus stable se dessine. Les dépôts décrits par des mots-clés et des résumés continuent de gagner du terrain, atteignant environ 43,8 % en 2016, puis se stabilisant autour de 40 % en 2023. Parallèlement, les dépôts décrits par des mots-clés ou des résumés représentent une part significative, oscillant entre 27 % et 32 % au cours des dernières années. Malgré ces améliorations, la part des dépôts sans description reste notable, bien qu'elle ait progressivement diminué, et se situe aux alentours de 27 % en 2023. Cette évolution des pratiques d'archivage dans les STM montre une maturation progressive des méthodes de dépôt.

Bien que plusieurs années ont été marquées par une faible attention portée à la description des notices, les dernières années ont vu une amélioration notable sur ce point, alors même que le nombre de dépôts annuels augmente. L'importance des métadonnées pour la découverte et la visibilité des travaux scientifiques commence à être mieux reconnue par les chercheurs mais certains défis subsistent pour améliorer l'efficacité de l'archivage pour la recherche. Ce constat, déjà posé au chapitre précédent, invite à reconsidérer les métadonnées comme des constructions sociotechniques (Zacklad, 2004) plutôt que comme de simples attributs techniques. Leur complétude, ou au contraire leur absence, témoigne de choix qui traduisent des arbitrages sur ce que les chercheurs rendent visible, traçable, voire évaluable. Cette perspective permet d'interpréter la qualité des dépôts non pas uniquement en termes de conformité aux standards, mais aussi comme révélatrice des rapports différenciés à l'archivage et plus généralement la science ouverte.

Figure 26

Évolution des niveaux de description des dépôts dans le temps en STM



La figure 27 montre l'évolution des niveaux de description des dépôts dans le temps, en sciences, techniques et médecine (STM), selon les logiques d'archivage et de communication scientifique directe (haut), ainsi que selon les logiques de référencement et de recensement (bas), pour la période allant de 2002 à 2023.

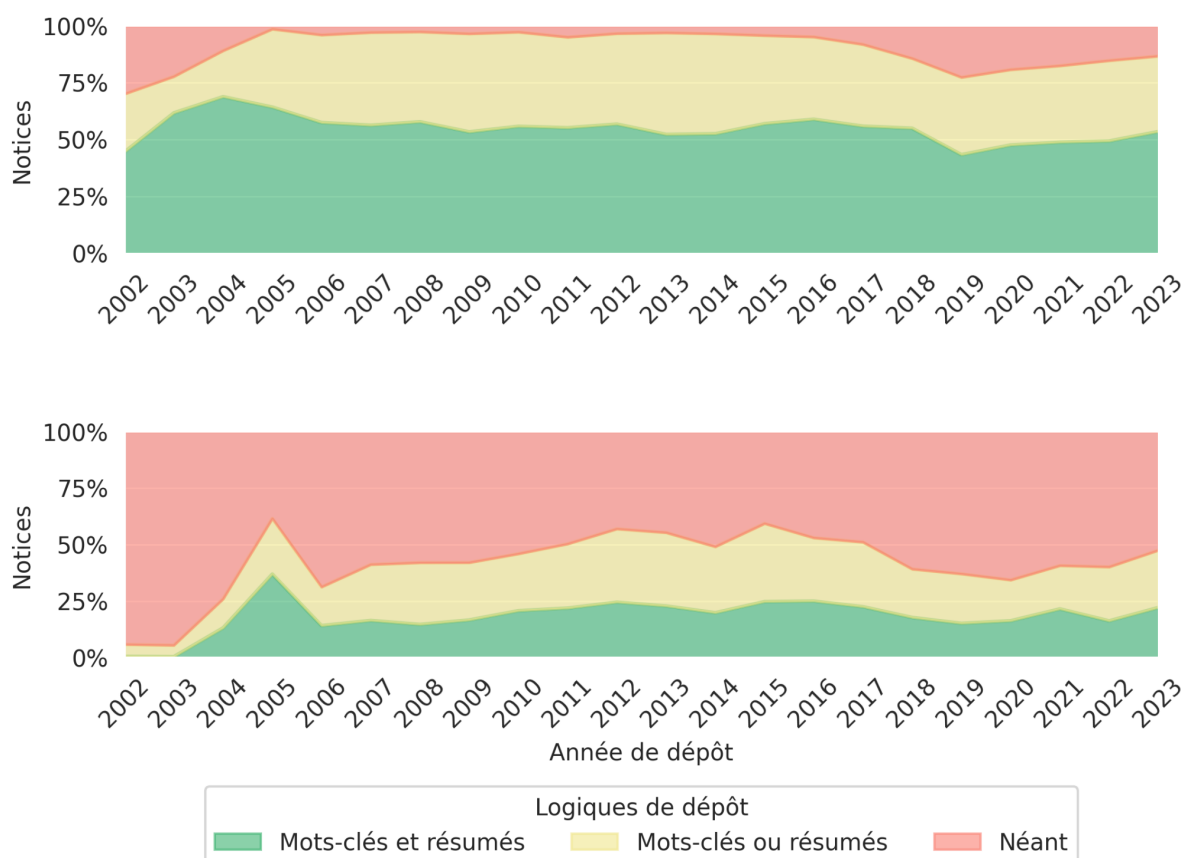
Pour les logiques d'archivage et de communication scientifique directe, on observe une stabilité relative des dépôts comportant à la fois des mots-clés et des résumés. Ces derniers représentent une part significative au début de la période, atteignant un pic autour de 60 % en 2004. Cette proportion tend à diminuer légèrement à partir de 2007, oscillant autour de 40 % jusqu'en 2023. Cette stabilité relative montre que, malgré une légère baisse, les pratiques de dépôt dans ces logiques sont caractérisées par une attention portée aux métadonnées descriptives. L'inclusion simultanée de mots-clés et de résumés facilite grandement la diffusion et la trouvabilité des travaux déposés. Les dépôts ne comportant ni mots-clés ni résumés, bien que relativement stables autour de 20 à 30 %, indiquent qu'une partie non négligeable des dépôts n'inclut toujours pas de descriptions complètes, ce qui pourrait limiter leur visibilité à long terme.

En revanche, pour les logiques de référencement et de recensement, la situation est nettement différente. Les dépôts ne comportant aucune métadonnée descriptive dominent largement tout au long de la période, culminant à plus de 60 % en 2004 et 2005, avant de diminuer progressivement jusqu'à environ 40 % en 2023. Ces dépôts sont typiques des pratiques de

recensement et de référencement, où l'objectif principal est la simple visibilité des travaux, sans réelle intention de diffusion. Les dépôts associant des mots-clés et des résumés restent faibles tout au long de la période, ne dépassant jamais 25 % des dépôts annuels à l'exception de 2004. Ce résultat indique une absence persistante de renseignement des métadonnées, particulièrement dans les logiques de recensement et de référencement, ce qui réduit la trouvabilité et par extension l'utilisation future de ces travaux. Les contributeurs qui suivent ces logiques s'inscrivent donc dans un registre de visibilité minimale.

Figure 27

Évolution des niveaux de description des dépôts dans le temps en STM déposées par selon des logiques d'archivage et de communication scientifique directe (haut) et selon des logiques de référencement et de recensement (bas)



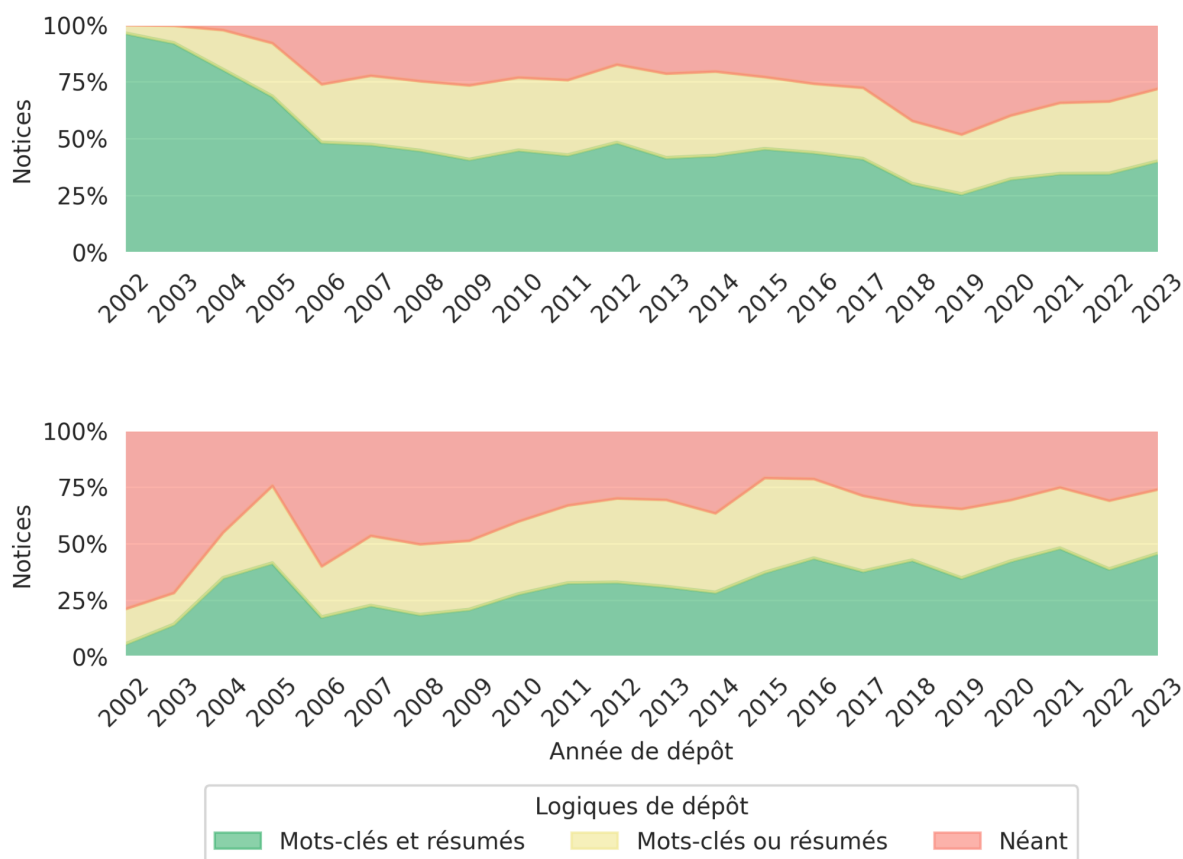
La figure 28 présente l'évolution des niveaux de description des dépôts dans le temps en STM, en distinguant les contributeurs en auto-archivage (haut) et les contributeurs intermédiaires (bas), sur la période allant de 2002 à 2023.

Pour les contributeurs en auto-archivage, on observe que la proportion des dépôts décrits par mots-clés et des résumés est dominante en début de période, atteignant environ 96 % en 2002. Toutefois, cette proportion diminue progressivement au fil des années pour se stabiliser autour de 40 % à partir de 2010. Cela montre une transition vers des dépôts moins bien décrits. Les dépôts décrits par des mots-clés ou des résumés augmentent légèrement, mais ne dépassent jamais une part importante du total, oscillant autour de 5 à 10 % durant la période étudiée. Enfin, bien que la proportion de dépôts non décrits augmente légèrement au fil du temps, elle reste importante mais globalement minoritaire pour les contributeurs en auto-archivage. Un pic notable est observé en 2019, où 48 % des dépôts ne comportaient pas de description, mais cette tendance s'atténue par la suite. Cette dynamique suggère que ces contributeurs, intrinsèquement majoritairement des chercheurs, cherchent à décrire au mieux leurs travaux dans le but d'optimiser la visibilité de ces derniers, même si cet effort diminue avec le temps. Cette diminution coïncide avec l'augmentation du nombre de dépôts annuels.

Le comportement des contributeurs intermédiaires diffère notablement. En 2002, plus de 79 % des dépôts réalisés par des intermédiaires ne comportaient ni mots-clés ni résumés. Cette proportion baisse progressivement pour atteindre environ 50 % en 2023. Cette tendance montre un effort croissant des intermédiaires pour enrichir les dépôts avec des métadonnées descriptives, bien que la majorité des dépôts restent encore sans description complète. La part des dépôts décrits par à la fois des mots-clés et des résumés est restée relativement faible tout au long de la période, atteignant à peine 20 % à partir de 2010. Cette différence avec les contributeurs en auto-archivage s'explique par la nature même des dépôts réalisés par des intermédiaires, souvent liés à des pratiques de recensement ou de référencement, qui ne nécessitent pas forcément une complétude des métadonnées pour atteindre leur objectif principal de visibilité institutionnelle. Toutefois, sur la fin de la période, on note que les deux niveaux de contribution atteignent des pratiques de description similaires. Cette convergence progressive peut s'expliquer par une standardisation des pratiques de dépôt, renforcée par des directives institutionnelles visant à améliorer la qualité des dépôts, quel que soit le type de contributeur. Les efforts pour uniformiser la qualité des métadonnées, même pour les dépôts effectués par des intermédiaires, pourraient être le reflet d'une prise de conscience collective de l'importance de la complétude des descriptions pour la trouvabilité et l'impact scientifique des travaux. Ainsi, malgré des trajectoires initialement divergentes, les contributeurs intermédiaires tendent progressivement vers des pratiques plus proches de celles des chercheurs en auto-archivage, réduisant ainsi l'écart qui caractérisait les premières années de la période observée.

Figure 28

Évolution des niveaux de description des dépôts dans le temps en STM déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)



La figure 29 présente l'évolution des niveaux de description des dépôts dans le temps en STM, en distinguant les contributeurs réguliers (haut) et les nouveaux contributeurs (bas) sur la période de 2004 à 2021.

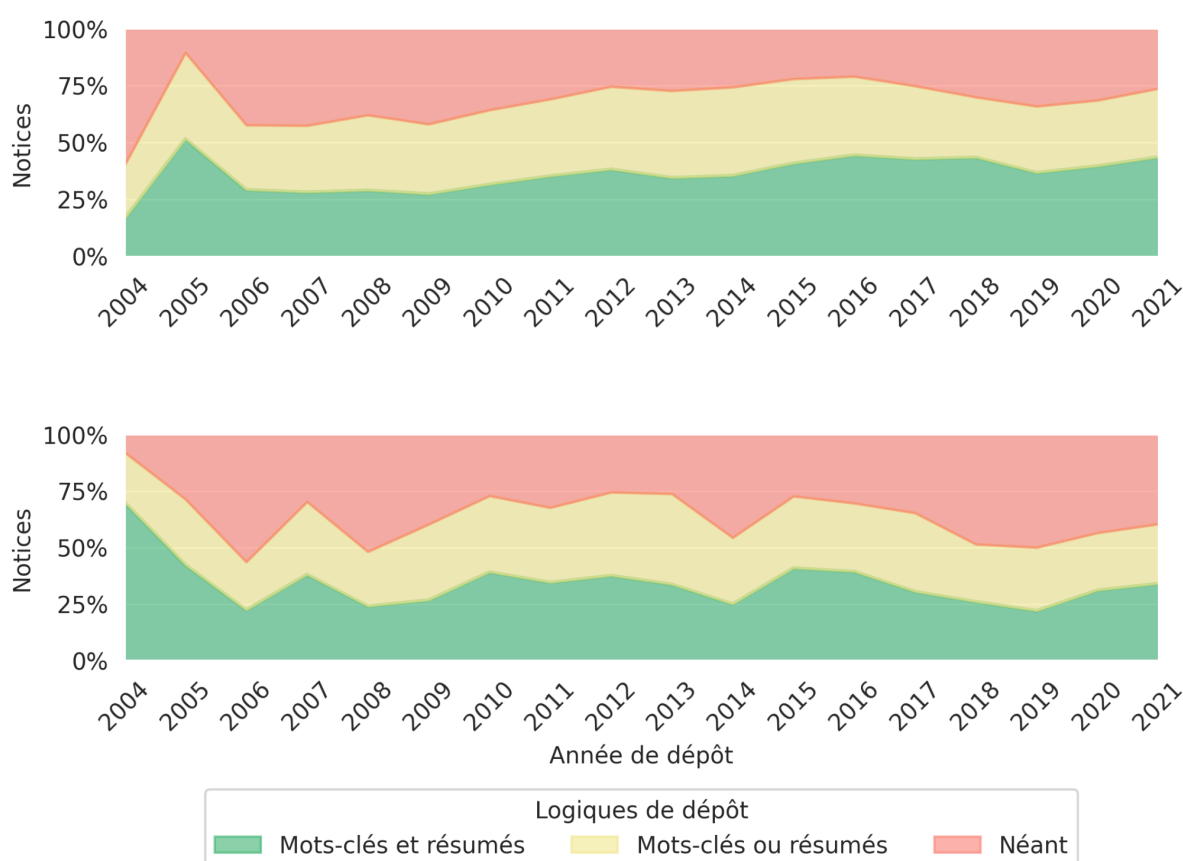
Pour les contributeurs réguliers, on observe une stabilisation des dépôts décrits par à la fois des mots-clés et des résumés autour de 40 % à partir de 2006, avec une légère augmentation progressive jusqu'en 2021. Cela indique que les contributeurs réguliers maintiennent un effort constant pour enrichir leurs dépôts avec des métadonnées complètes, même si cet effort fluctue légèrement au fil des ans. En revanche, la part des dépôts non décrits connaît un déclin progressif, tombant de près de 50 % en 2005 à environ 20 % en 2021, ce qui montre une amélioration du niveau global de description des dépôts avec le temps.

La dynamique des nouveaux contributeurs, en revanche, est bien plus irrégulière tout au long de la période observée. En 2006, la part des dépôts sans description complète atteint un pic de près

de 75 %, mais diminue ensuite de manière inégale pour atteindre environ 40 % en 2021. Les dépôts décrits par à la fois des mots-clés et des résumés montrent une augmentation à partir de 2008, mais cette croissance est également irrégulière, culminant à environ 30 % en 2021. Enfin, la part des dépôts décrits seulement par des mots-clés ou des résumés reste assez instable, variant de 20 % à 30 % sans réelle tendance claire. Contrairement à la stabilité relative observée chez les contributeurs réguliers, les fluctuations marquées chez les nouveaux contributeurs rendent difficile l'identification d'une tendance claire sur l'ensemble de la période. Néanmoins, nous observons que le niveau de description global reste moins élevé pour ces derniers que pour les contributeurs réguliers. Ce décalage reflète probablement la moindre expérience, et par conséquent, le besoin de formation pour ce type de contributeurs.

Figure 29

Évolution des niveaux de description des dépôts dans le temps en STM déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)



3.2.2. *Discussion intermédiaire*

L'analyse détaillée de l'évolution des pratiques d'archivage dans les domaines des sciences, techniques et médecine (STM) met en lumière les dynamiques complexes qui caractérisent la contribution et la gestion des dépôts au fil du temps. En identifiant des phases distinctes dans l'adoption des pratiques d'archivage, cette cartographie révèle une progression constante vers une maturité institutionnelle et collective en matière de gestion et de diffusion des résultats scientifiques.

Les données montrent une diversification des logiques de dépôt, marquée par une transition d'une dominance initiale de la communication scientifique directe à une dispersion progressive vers des logiques de recensement, de référencement et d'archivage. Cette évolution témoigne de l'adaptation des communautés scientifiques aux enjeux croissants de pérennité, de traçabilité et de visibilité des savoirs. Cependant, des disparités subsistent entre les différents types de contributeurs et niveaux de contribution. Les contributeurs réguliers et en auto-archivage privilégient une description plus détaillée des dépôts, tandis que les nouveaux contributeurs et les contributeurs intermédiaires tendent à adopter des pratiques moins complètes, bien qu'une convergence progressive se dessine en fin de période.

Enfin, l'évolution des niveaux de description des dépôts souligne des efforts croissants pour améliorer la complétion des métadonnées, mais met également en exergue des défis persistants. Ces observations peuvent être lues à travers le prisme des logiques de prescription et d'appropriation (Jouët, 2000) : si certains chercheurs décrivent rigoureusement leurs dépôts, c'est souvent en raison d'une politique locale incitative ou d'un dispositif accompagnateur. A contrario, l'absence de métadonnées peut révéler un acte « obligé », réalisé pour satisfaire une exigence administrative sans réelle appropriation du geste d'archivage. Cela souligne l'importance de l'environnement sociotechnique dans lequel s'inscrivent les dépôts, et renforce la pertinence des dispositifs d'accompagnement comme SoVisu. Une part non négligeable des dépôts reste encore dépourvue de description complète, compromettant leur découvrabilité et leur impact scientifique. L'importance d'une description complète est soulignée par plusieurs travaux qui montrent que des métadonnées riches et précises facilitent la recherche et l'indexation des documents dans les moteurs de recherche (Gallezot et al., 2003). Ces constats appellent à une sensibilisation renforcée des différents acteurs, auteurs, institutions, administrateurs d'archives, pour assurer une gestion optimale des archives scientifiques. Il serait nécessaire de mettre en place des formations et un accompagnement personnalisé pour guider les chercheurs dans le processus

de dépôt. Dans cette perspective, il est utile de rappeler ce que signifie véritablement « s'approprier » un objet technique. Comme le formule Josiane Jouët :

L'appropriation est un processus. Elle est l'acte de se constituer un "soi", un projet, une histoire. Elle passe par la sélection et la réinterprétation des objets, par leur mise en sens et en cohérence dans une histoire personnelle ou collective. Elle engage une dynamique de transformation du sujet, des objets techniques et de leurs usages.

(Jouët, 2000)

Cette définition est idéale pour mettre en valeur l'idée que l'appropriation ne va pas de soi : elle nécessite un environnement favorable pour transformer une contrainte administrative, comme peut l'être l'acte de dépôt, en acte significatif. Autrement dit, le dépôt n'est pleinement efficace que lorsqu'il est intégré dans une dynamique personnelle, institutionnelle et technique cohérente.

En somme, cette cartographie met en évidence les transformations structurelles et culturelles au sein des disciplines STM, reflétant une appropriation croissante des outils d'archivage ouverts. Ces résultats offrent une base solide pour approfondir les réflexions sur les stratégies futures de gestion des dépôts.

Ces tendances doivent être mises en perspective avec les évolutions de la plateforme HAL à cette période, notamment les flux de dépôt automatisés par certaines revues ou la création de portails universitaires. Ces transformations techniques peuvent expliquer en partie les progressions quantitatives observées.

3.2.3. *Les sciences humaines et sociales*

Nous poursuivons notre exploration en étudiant les sciences humaines et sociales. Cette analyse reprend le plan développé précédemment pour les sciences, techniques et médecine tout en soulignant les spécificités et différences notables dans les pratiques d'archivage. Les sciences humaines et sociales avec leurs approches épistémologiques distinctes présentent des dynamiques de contribution qui diffèrent significativement de celles observées dans les disciplines plus « dures ». Alors que la précédente partie met en évidence des pratiques guidées par des paradigmes de diffusion plus structurés, cette partie révèle comment les pratiques d'archivage en sciences humaines et sociales sont influencées par une diversité d'approches et de contextes.

Les figures 30 et 31 présentent respectivement l'évolution du nombre de dépôts et du nombre de contributeurs uniques dans le domaine des sciences humaines et sociales (SHS) de 2002 à 2023.

Les données issues de ces graphiques révèlent plusieurs tendances marquantes dans l'adoption et l'intensification des pratiques d'archivage au sein de ces disciplines. Comme pour les sciences, techniques et médecine, nous observons ici trois phases distinctes.

Au début de la période, de 2002 à 2008, le nombre de dépôts progresse à un rythme relativement faible. Le nombre de contributeurs uniques suit une trajectoire similaire, marquant une phase d'adoption initiale des pratiques d'archivage, avec une sensibilisation encore en développement au sein de la communauté des chercheurs en SHS.

La deuxième phase, de 2009 à 2016, se distingue par une croissance modérée et irrégulière du nombre de dépôts, accompagnée d'une augmentation graduelle, mais plus lente, du nombre de contributeurs. Cette période est marquée par une consolidation des pratiques d'archivage, avec une base de contributeurs en expansion qui commence à s'intégrer plus systématiquement dans les processus d'auto-archivage.

À partir de 2017, une inflexion nette apparaît dans la dynamique de dépôt et de participation, marquée par une accélération significative de la croissance, tant en termes de nombre de dépôts que de contributeurs uniques. Cette période est caractérisée par une montée en puissance rapide et soutenue, où les chiffres témoignent d'une adoption importante de l'archivage. Ces évolutions coïncident avec les initiatives gouvernementales mises en œuvre en faveur de la science ouverte. Le nombre croissant de dépôts et de contributeurs indique que de plus en plus de chercheurs s'engagent dans le processus d'archivage. Cette phase d'expansion exponentielle reflète également une maturation des pratiques d'archivage, où l'archivage ouvert s'affirme de plus en plus comme une voie centrale pour la diffusion du savoir scientifique.

Figure 30

Évolution de la quantité de dépôts dans le temps en SHS

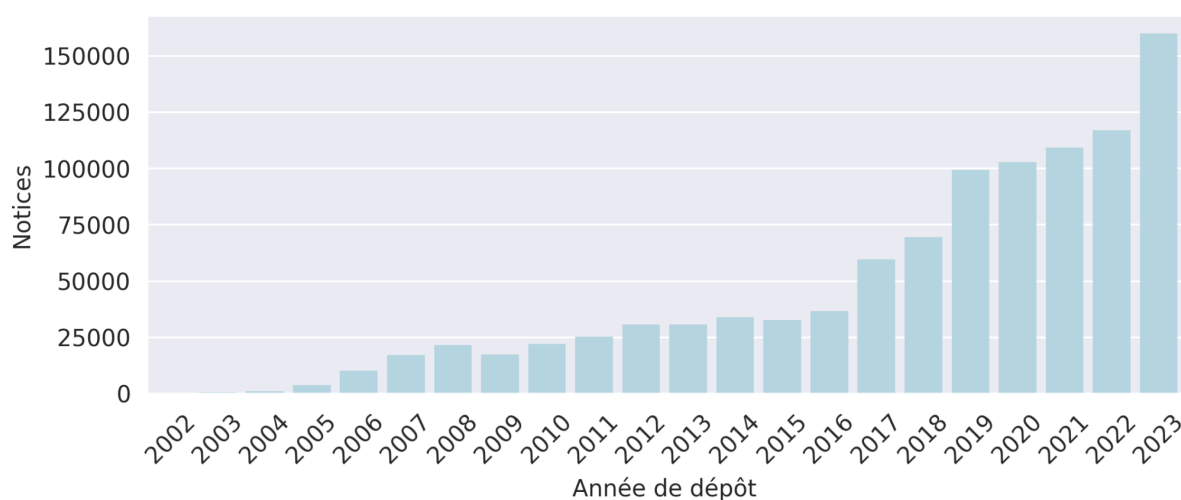
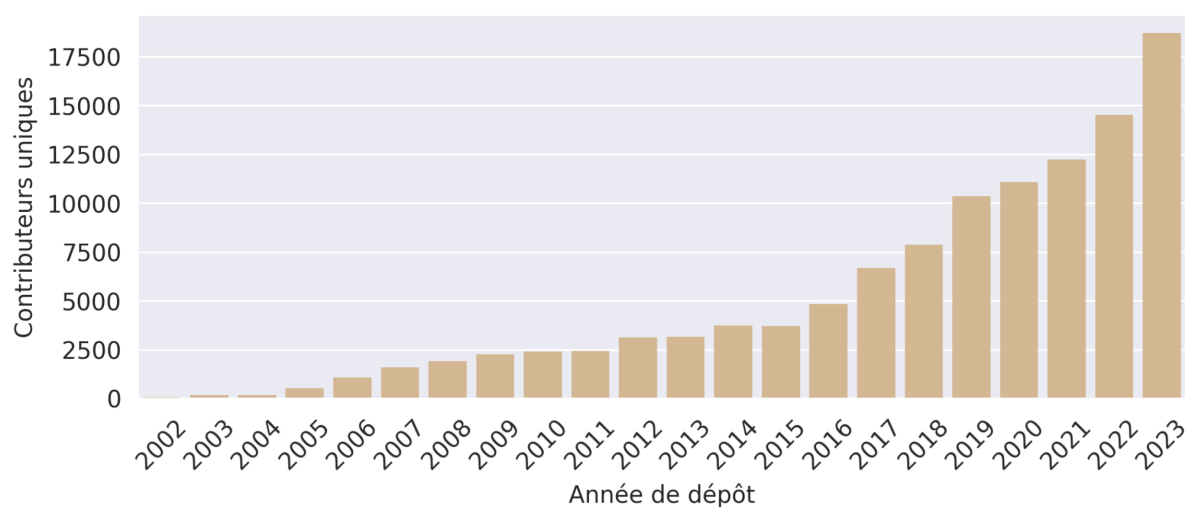


Figure 31

Évolution du nombre de contributeurs uniques dans le temps en SHS



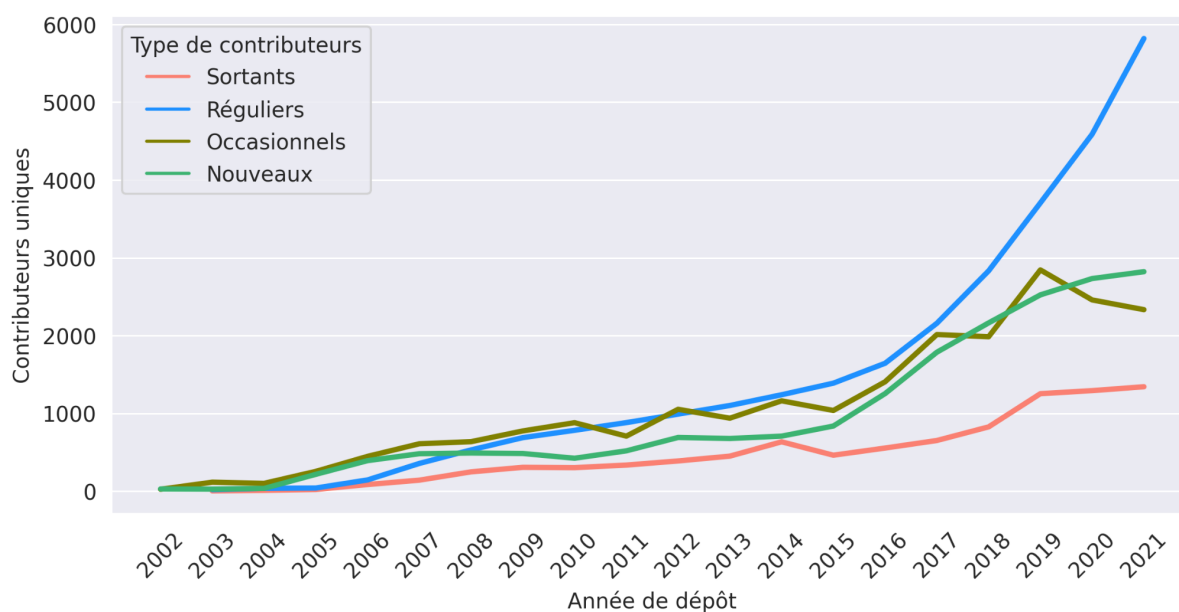
La figure 32 illustre l'évolution des types de contributeurs dans le domaine des sciences humaines et sociales (SHS) de 2002 à 2021. Pour rappel, ce graphique distingue quatre types de contributeurs : les réguliers (ligne bleue), les occasionnels (ligne verte), les nouveaux (ligne jaune), et les sortants (ligne rouge).

L'analyse de ces courbes permet de mettre en lumière plusieurs dynamiques importantes dans la participation aux pratiques d'archivage. Au début de la période, entre 2002 et 2008, le nombre de contributeurs reste relativement faible quel que soit le type, témoignant d'une adoption

progressive des pratiques d'archivage au sein de la communauté des chercheurs en SHS. Les contributeurs nouveaux et occasionnels dominent légèrement en nombre, indiquant une phase d'expérimentation où les chercheurs commencent à s'engager dans l'archivage sans encore adopter ces pratiques de manière régulière. À partir de 2009, on observe une augmentation progressive du nombre de contributeurs nouveaux et réguliers, qui commence à croître de manière plus soutenue. Cette période, marquée par une consolidation des pratiques d'archivage, voit une expansion de la base de contributeurs qui s'intègrent progressivement dans le processus d'auto-archivage. Les contributeurs occasionnels suivent une tendance similaire, mais avec une légère fluctuation, tandis que le nombre de contributeurs sortants reste relativement stable. À partir de 2017, la dynamique change radicalement. Le nombre de contributeurs réguliers connaît une croissance exponentielle, atteignant son apogée en 2021 avec un pic significatif. Cette augmentation témoigne de l'adoption massive des pratiques d'archivage par un nombre croissant de chercheurs, soutenue par les initiatives gouvernementales en faveur de la science ouverte. Les contributeurs nouveaux continuent également d'augmenter, bien qu'à un rythme plus modéré à partir de 2019, suggérant une certaine saturation ou stabilisation dans l'intégration de nouveaux membres. Les contributeurs occasionnels suivent une tendance similaire, avec une légère baisse après 2018, tandis que les contributeurs sortants montrent une augmentation modeste, indiquant un renouvellement partiel au sein de la communauté.

Figure 32

Évolution des types de contributeurs dans le temps en SHS



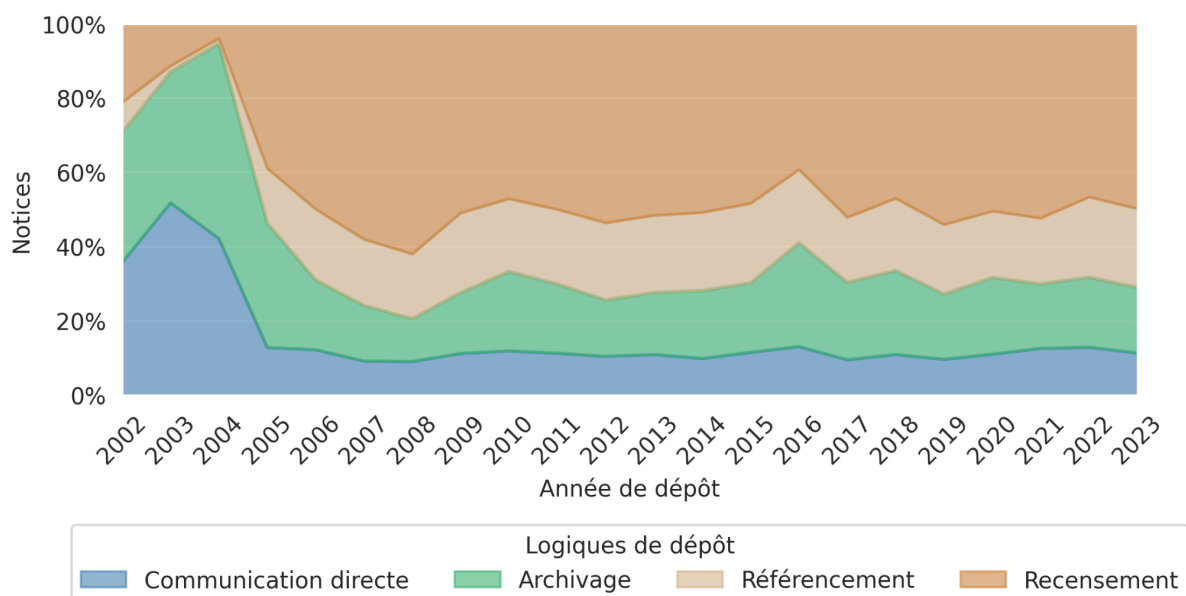
La figure 33 illustre l'évolution des logiques de dépôt dans le domaine des sciences humaines et sociales (SHS) entre 2002 et 2023. Ce graphique à aires empilées, permet de visualiser comment la répartition des différentes logiques de dépôt a changé au fil du temps, reflétant des transformations dans les pratiques d'archivage au sein de cette communauté scientifique.

Au début de la période, la communication scientifique directe occupe une place prépondérante, avec un pic notable en 2003 où elle représente environ 52 % des dépôts. Cette domination initiale reflète une volonté forte de diffuser rapidement les résultats de recherche, avec un accent mis sur l'immédiateté de la communication scientifique. Cependant, cette tendance s'inverse dès 2004, où la communication directe commence à perdre du terrain de manière significative. À partir de 2005, un changement important se produit : la part de la communication directe diminue fortement, tandis que les logiques de recensement et de référencement gagnent en importance. Le recensement, en particulier, prend une position dominante à partir de 2006, représentant environ 50 % des dépôts. Cette période marque un tournant vers des pratiques d'archivage davantage axées sur la visibilité et la documentation des recherches passées, même sans la présence d'un texte intégral. Le référencement suit également cette tendance, devenant une composante clé de la stratégie de dépôt, en particulier à partir de la seconde moitié des années 2000. À partir de 2010, les pratiques d'archivage se stabilisent et se diversifient. L'archivage, qui consiste à déposer des travaux plus anciens accompagnés d'un texte intégral, commence à jouer un rôle de plus en plus important, atteignant environ 17 à 20 % des dépôts en 2023. Pendant cette même période, la communication scientifique directe, après avoir chuté dans les années précédentes, stabilise sa part autour de 10 à 13 %. Le recensement continue de dominer, mais de manière plus équilibrée, avec des pourcentages qui oscillent autour de 50 % dans les dernières années. Le référencement, quant à lui, reste stable, représentant environ 21 % des dépôts en 2023.

Cette évolution des pratiques d'archivage dans les SHS révèle une transition progressive, passant d'une prédominance initiale de la diffusion immédiate des résultats à une répartition plus diversifiée des logiques de dépôt. Si le recensement et le référencement ont gagné en importance, notamment en raison de leur capacité à améliorer la visibilité des travaux, cette évolution ne doit pas masquer les défis que posent ces pratiques. En effet, la montée du recensement et du référencement, bien qu'elle puisse indiquer une volonté accrue de documenter les recherches passées, peut aussi être perçue comme un signe d'un certain affaiblissement de la fonction première d'une archive ouverte, qui est de diffuser largement et librement les connaissances scientifiques.

Figure 33

Évolution des logiques de dépôt dans le temps en SHS



La figure 34 illustre l'évolution des logiques de dépôt dans les domaines sciences humaines et sociales (SHS), avec une distinction entre les contributeurs réguliers (haut) et les nouveaux contributeurs (bas), sur la période allant de 2002 à 2023.

Pour les contributeurs en auto-archivage, la logique de communication scientifique directe joue un rôle majeur au début de la période, atteignant 64,6 % en 2003, puis diminuant de façon marquée à partir de 2005 pour se stabiliser autour de 11-16 % entre 2016 et 2023. En parallèle, la logique d'archivage, qui correspond aux dépôts différés mais avec document associé, fluctue tout au long de la période, atteignant environ 28 % en 2016 avant de redescendre à environ 18 % en 2023. Une part croissante des dépôts suit une logique de recensement (sans document associé), passant de 22 % en 2002 à 46 % en 2023. Cette tendance montre que les contributeurs en auto-archivage se concentrent de plus en plus sur la visibilité des recherches passées, sans fournir systématiquement un document en texte intégral. La logique de référencement, quant à elle, varie légèrement au fil des ans, atteignant environ 24 % en 2023.

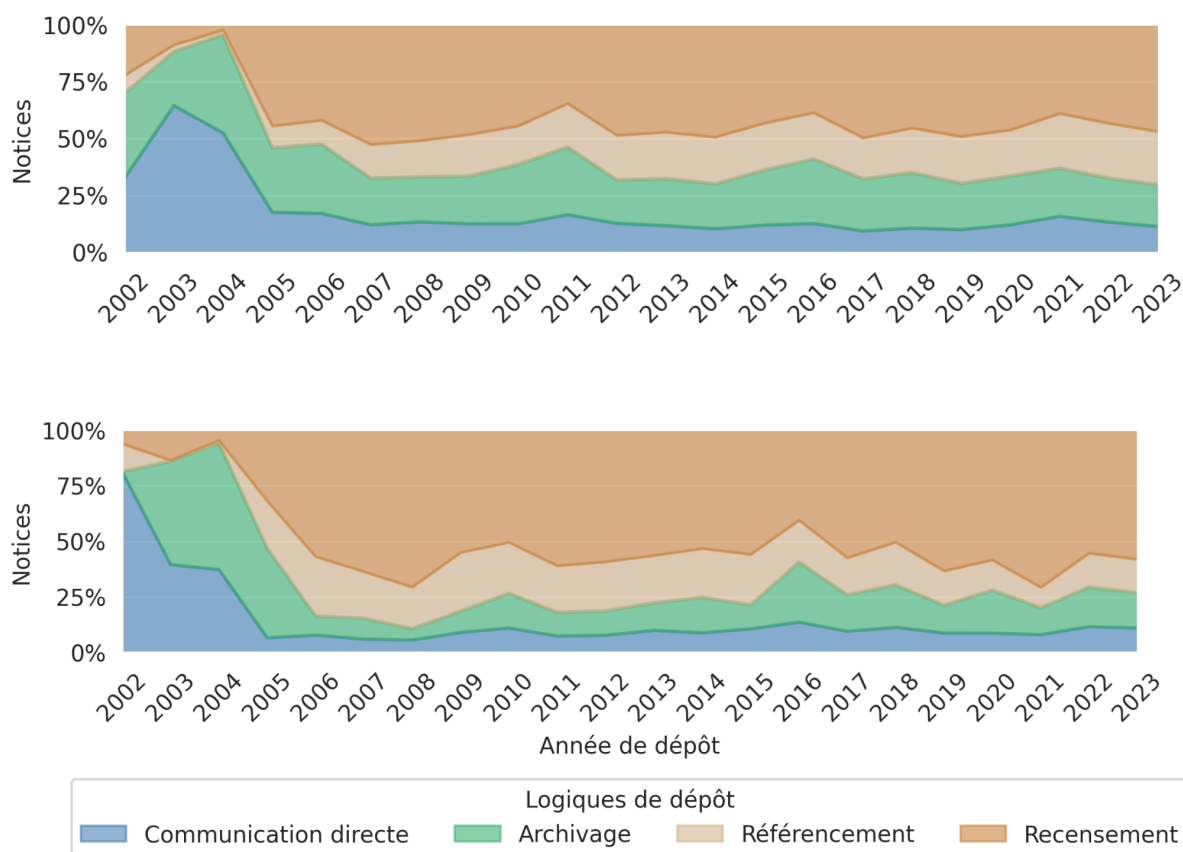
Pour les contributeurs intermédiaires, la logique de communication scientifique directe est moins dominante. En 2003, elle atteint 39,5 %, mais décline rapidement après 2005, pour se stabiliser autour de 8-11 % entre 2016 et 2023. Contrairement aux contributeurs en auto-archivage, ces contributeurs suivent plus souvent une logique de recensement. Cette dernière logique est particulièrement importante, représentant plus de 70 % des dépôts en 2021, bien qu'elle

redescende à 58 % en 2023. Cette préférence pour le recensement sans texte intégral montre une différence significative avec les contributeurs en auto-archivage. De même, l'archivage connaît une évolution contrastée, avec des pics comme en 2016 (27 %) avant de redescendre à environ 16 % en 2023. La logique de référencement reste relativement stable, représentant environ 9 à 26 % des dépôts tout au long de la période, mais tend à diminuer vers la fin, atteignant environ 15 % en 2023.

En comparant les deux niveaux de contribution, on observe que les contributeurs en auto-archivage privilégient davantage la communication scientifique directe que les contributeurs intermédiaires, surtout au début de la période, bien que cette tendance diminue au fil du temps. Les contributeurs intermédiaires, quant à eux, s'appuient davantage sur la logique de recensement, avec des niveaux plus élevés de dépôts sans document associé tout au long de la période. Les contributeurs en auto-archivage tendent à favoriser la diffusion immédiate des documents dans une logique de communication scientifique directe, tandis que les contributeurs intermédiaires optent davantage pour des dépôts sans texte intégral et priorisent la visibilité des recherches antérieures. Selon nous, cette absence de fichier associé s'explique probablement par le fait que l'intermédiaire ne dispose pas cet élément.

Figure 34

Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)



La figure 35 illustre l'évolution des logiques de dépôt dans les domaines sciences humaines et sociales (SHS), avec une distinction entre les contributeurs réguliers (haut) et les nouveaux contributeurs (bas), sur la période allant de 2004 à 2021.

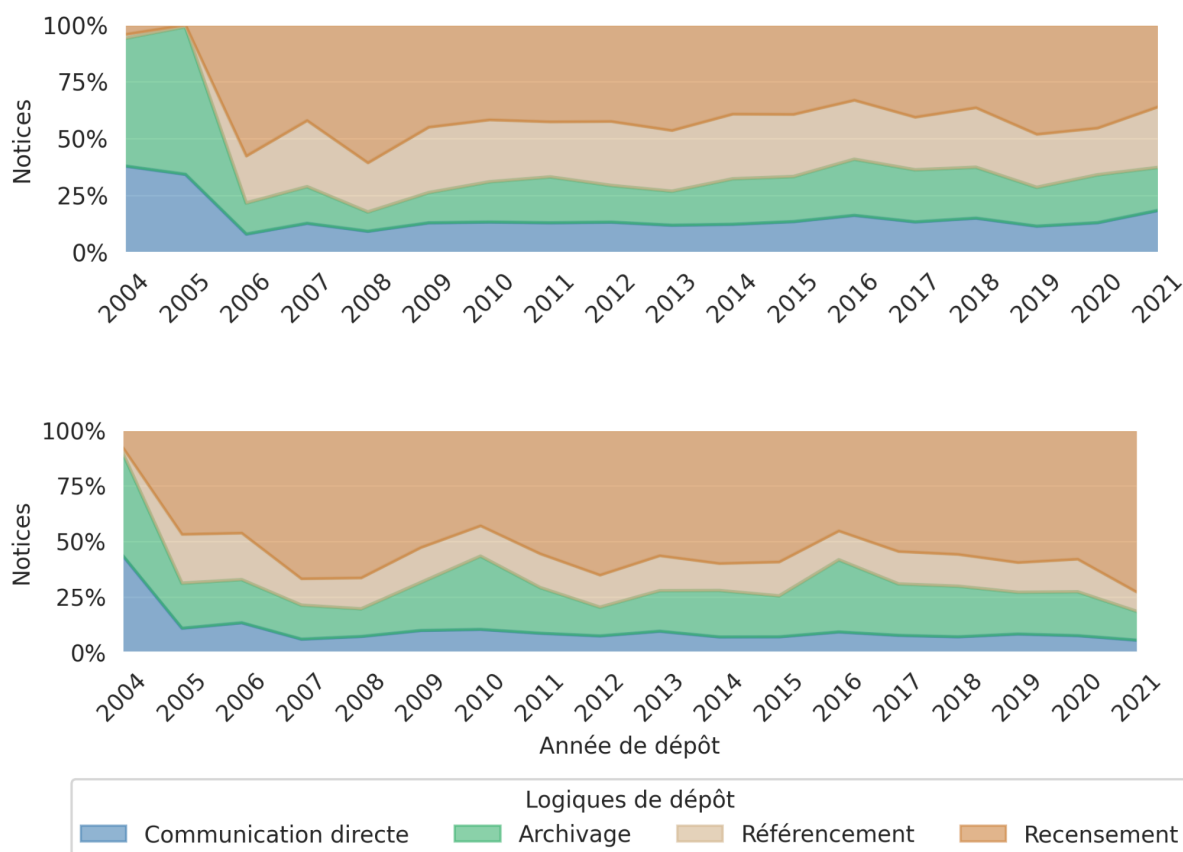
Pour les contributeurs réguliers, la logique de communication scientifique directe domine au début de la période, atteignant un pic de 64,9 % en 2005, puis décroît progressivement pour se stabiliser autour de 18-22 % dans les dernières années. Cette tendance marque une diminution du recours à la diffusion immédiate des documents. En parallèle, la logique de l'archivage commence à des niveaux modestes en 2004 (56 %) et suit un schéma irrégulier, atteignant environ 18,9 % en 2021. La logique de recensement prend de l'ampleur après 2006, notamment en 2013 où il atteint 46,3 %, avant de retomber autour de 35-45 % dans les dernières années de la période. Quant à la logique de référencement, il reste relativement stable, oscillant entre 24 % et 29 % tout au long de la période.

Pour les nouveaux contributeurs, la dynamique est assez différente. La logique de communication scientifique directe montre une importance variable : elle atteint un sommet de 46 % en 2004, mais décline ensuite jusqu'à environ 12,9 % en 2021. Contrairement aux contributeurs réguliers, les nouveaux contributeurs déposent plus fréquemment selon une logique de recensement, qui domine largement dans la plupart des années de la période, culminant à 73 % en 2021. Cette logique semble être la plus courante pour les nouveaux contributeurs, soulignant une tendance à déposer des notices sans texte associé. La logique d'archivage joue un rôle mineur, fluctuant entre 5 % et 10 % pendant la majeure partie de la période, avec un pic temporaire à 32,4 % en 2016. La logique de référencement, quant à elle, reste faible tout au long de la période, ne dépassant jamais 22 % et atteignant son plus bas niveau en 2021 (8,5 %).

En comparant les deux types de contribution, on observe que les contributeurs réguliers ont un recours plus stable et soutenu aux logiques de communication scientifique directe et d'archivage, ce qui reflète une approche plus régulière et organisée de la diffusion et de la conservation des documents. À l'inverse, les nouveaux contributeurs favorisent davantage la logique de recensement, ce qui pourrait indiquer une préférence pour la visibilité des recherches sans diffuser les documents associés. Le référencement occupe une place mineure dans les deux groupes, mais il est légèrement plus important chez les contributeurs réguliers.

Figure 35

Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)



La figure 36 retrace l'évolution des niveaux de description des dépôts dans le domaine des sciences humaines et sociales (SHS) de 2002 à 2023, illustrant les changements progressifs dans les pratiques d'archivage au fil du temps. Le graphique distingue trois catégories de description : les dépôts accompagnés de mots-clés et de résumés (en vert), ceux avec mots-clés ou résumés (en jaune), et les dépôts sans description (néant, en rouge).

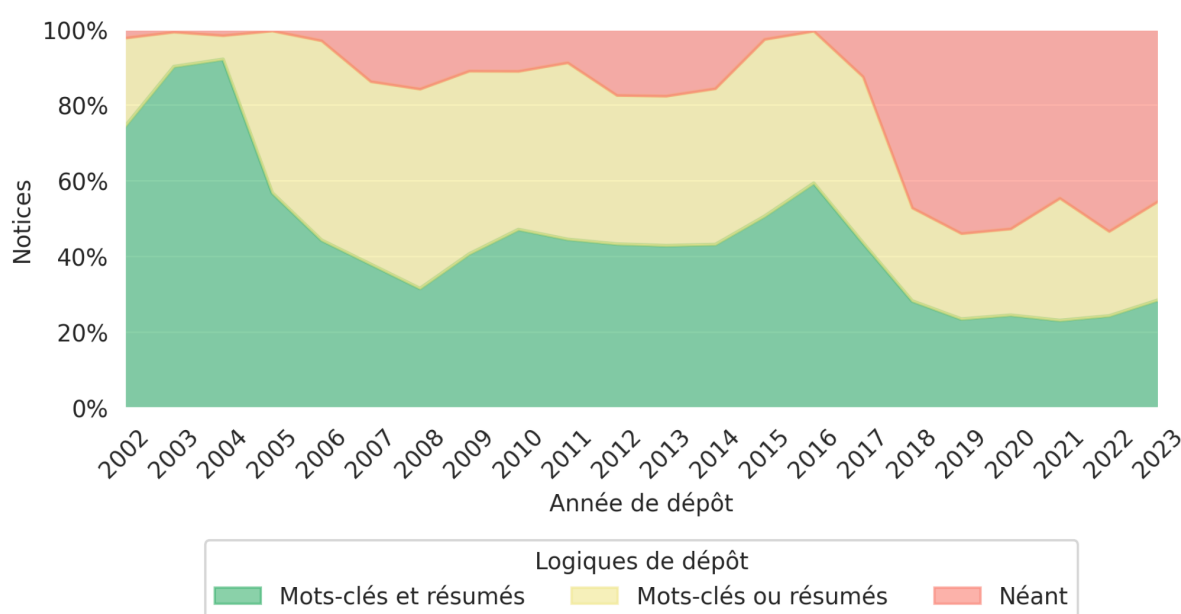
Au début des années 2000, les pratiques d'archivage dans les SHS sont marquées par une forte présence de métadonnées complètes, avec une nette majorité des dépôts accompagnés de mots-clés et de résumés. Cette tendance culmine en 2004, où cette catégorie représente plus de 92 % des dépôts. Cependant, cette période de rigueur initiale cède rapidement le pas à une phase de rééquilibrage. Dès 2005, et plus encore à partir de 2006, on observe une diminution progressive de la part des dépôts avec des métadonnées complètes, au profit des dépôts décrits seulement par des mots-clés ou des résumés, qui deviennent de plus en plus fréquents. Cette fluctuation se poursuit jusqu'en 2016, période durant laquelle les pratiques oscillent sans se stabiliser. À partir de

2017 un changement notable se produit dans la structure des dépôts conjointement avec l'augmentation des dépôts. La proportion de dépôts sans description commence à augmenter de manière significative, atteignant des sommets en 2019 et 2022, où plus de la moitié des dépôts ne sont plus décrits. Ce retour en force des dépôts sans description contraste avec les années précédentes, et soulève des questions quant à l'efficacité des pratiques d'archivage en matière de découverte et de visibilité des travaux. En parallèle, les dépôts décrits par des mots-clés et des résumés continuent de décliner, représentant environ 24 % des dépôts en 2022, tandis que ceux accompagnés de mots-clés ou de résumés oscillent autour de 22 % à 26 %.

Cette tendance à la diminution des métadonnées complètes au profit de pratiques moins documentées pourrait refléter un relâchement des efforts pour fournir des descriptions détaillées, malgré une augmentation continue du nombre total de dépôts. Cette évolution des pratiques d'archivage dans les SHS montre une transition complexe. Alors que les premières années étaient caractérisées par une attention soutenue à la description des dépôts, les dernières années révèlent un recul préoccupant dans cette pratique, à un moment où l'archivage devrait, au contraire, renforcer ses exigences en matière de métadonnées pour garantir la visibilité et l'accessibilité des travaux scientifiques. L'importance des métadonnées, bien que reconnue, semble avoir été quelque peu relativisée, nécessitant une remise en question des stratégies d'archivage pour répondre aux défis actuels.

Figure 36

Évolution des niveaux de description des dépôts dans le temps en SHS



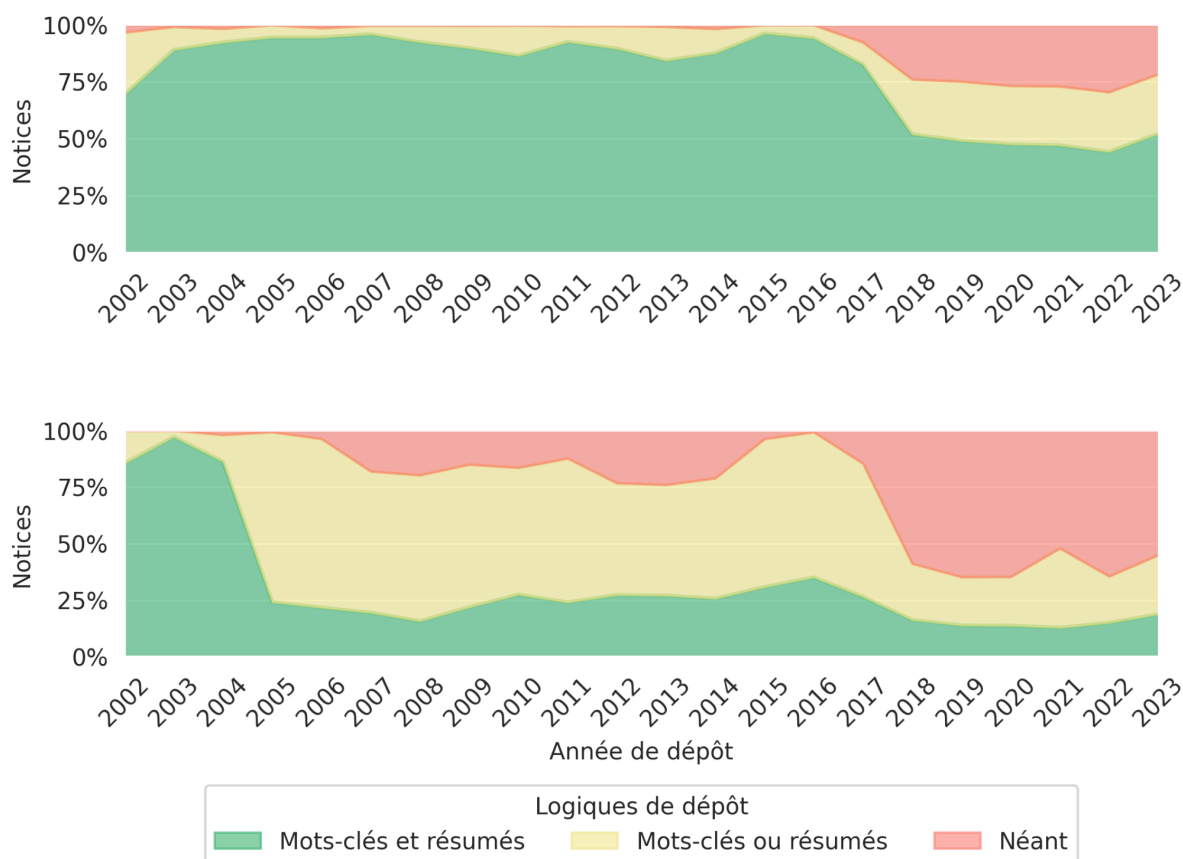
La figure 37 présente l'évolution des niveaux de description des dépôts dans le temps, en sciences humaines et sociales (SHS), selon les logiques d'archivage et de communication scientifique directe (haut) et les logiques de référencement et de recensement (bas), pour la période allant de 2002 à 2023.

Pour les logiques d'archivage et de communication scientifique directe, on observe qu'une large majorité des dépôts est accompagnée à la fois de mots-clés et de résumés au début de la période. En 2002, environ 70 % des dépôts sont décrits par ces métadonnées, et cette proportion augmente significativement pour atteindre près de 90 % en 2003. Ce niveau de description reste relativement stable jusqu'en 2018, où il subit une chute brutale, perdant plus de 40 points en l'espace d'une seule année. Le pourcentage des dépôts comportant à la fois des mots-clés et des résumés continue de diminuer progressivement, atteignant un minimum autour de 35 % en 2020, avant de remonter légèrement à environ 40 % en 2023. Cette évolution semble coïncider avec l'augmentation du nombre de dépôts annuels. Parallèlement, la proportion de dépôts sans description reste faible jusqu'en 2018, avant d'augmenter significativement pour atteindre environ 25 % en 2020.

Les dépôts effectués selon des logiques de référencement et de recensement suivent une tendance similaire, bien que de manière moins marquée. Au début de la période, la proportion de dépôts décrits par des mots-clés ou des résumés est relativement élevée. En 2002, environ 85 % des dépôts sont accompagnés de métadonnées descriptives complètes. Toutefois, cette proportion commence à diminuer rapidement dès 2004. À partir de cette année, la proportion de dépôts décrits uniquement par des mots-clés ou des résumés devient majoritaire jusqu'en 2018, atteignant près de 65 % en 2016. Cependant, cette proportion connaît une chute similaire à celle des dépôts effectués dans des logiques d'archivage et de communication scientifique directe. Cette dynamique traduit une moindre priorité accordée à la complétude des métadonnées dans les pratiques de référencement et de recensement, probablement en raison de leur objectif centré sur la visibilité institutionnelle plutôt que sur une diffusion exhaustive des contenus.

Figure 37

Évolution des niveaux de description des dépôts dans le temps en SHS déposées par selon des logiques d'archivage et de communication scientifique directe (haut) et selon des logiques de référencement et de recensement (bas)



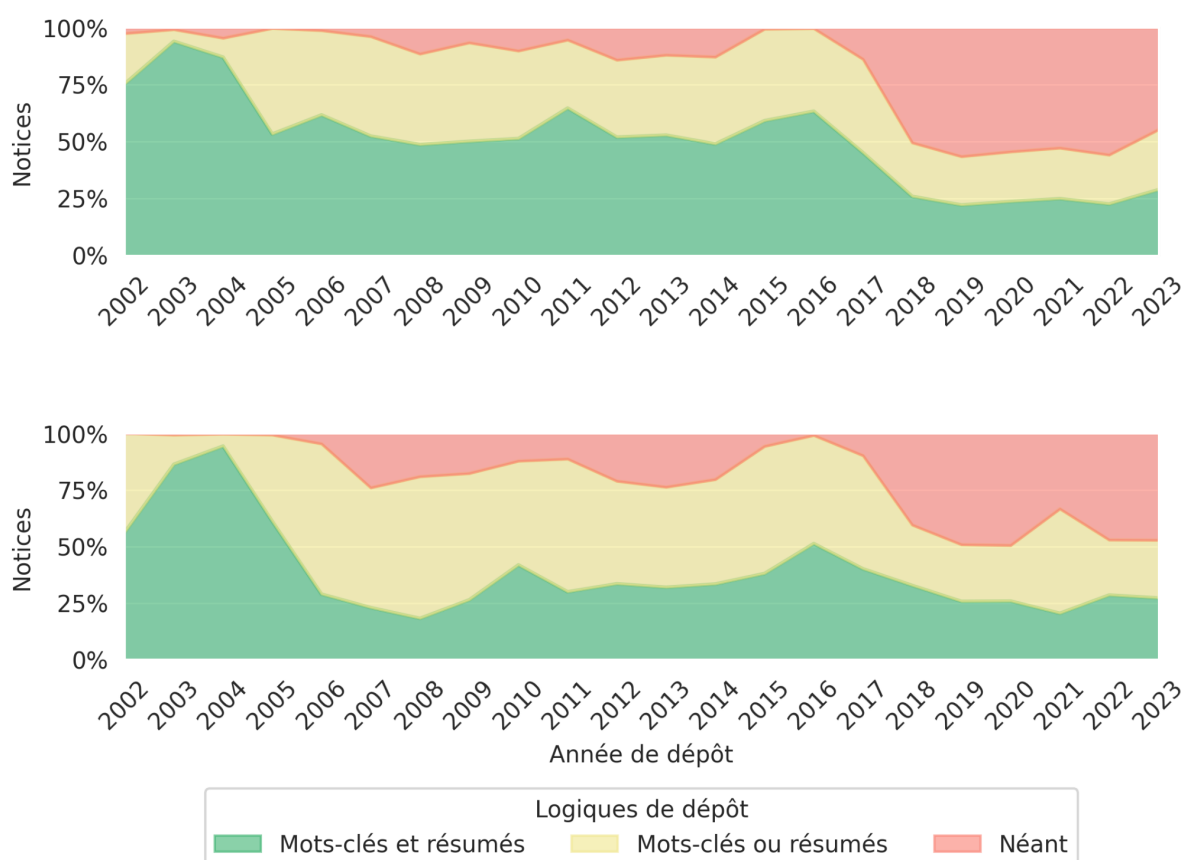
La figure 38 présente l'évolution des niveaux de description des dépôts dans le temps en sciences humaines et sociales (SHS), en distinguant les contributeurs en auto-archivage (haut) et les contributeurs intermédiaires (bas) sur la période allant de 2002 à 2023.

Pour les contributeurs en auto-archivage, on observe qu'au début de la période, une majorité des dépôts est décrite de manière complète, avec à la fois des mots-clés et des résumés. En 2002, environ 75 % des dépôts sont accompagnés de ces métadonnées, et cette proportion augmente significativement pour atteindre plus de 90 % en 2003. Cette proportion tombe aux alentours des 60 % dès 2005. Après une certaine stabilité jusqu'en 2016, cette proportion subit une chute notable à partir de 2018, pour atteindre un minimum autour de 35 % en 2020. Cette baisse coïncide, là encore, avec l'augmentation du nombre de dépôts annuels. Quant aux dépôts sans description, logiquement, leur part reste relativement faible en début de période, avant de connaître une hausse marquée dès 2018, atteignant environ 25 % en 2020.

En ce qui concerne les contributeurs intermédiaires, à l'image de la figure précédente, nous retrouvons également une tendance similaire mais de moindre degré. Les niveaux de description complets et aussi partiels, à savoir les dépôts décrits par des résumés et des mots-clés et des résumés ou des mots-clés suivent quasiment les mêmes variations que pour les dépôts effectués par des contributeurs en auto-archivage, y compris pour la chute soudaine de la description des dépôts dès 2018. La particularité des dépôts effectués par des contributeurs intermédiaires réside dans la proportion plus élevée de dépôts décrits par des mots-clés ou des résumés que celle des dépôts décrits de manière complète chez les contributeurs suivant les logiques exposées ci-dessus.

Figure 38

Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs en auto-archivage (haut) et par des contributeurs intermédiaires (bas)



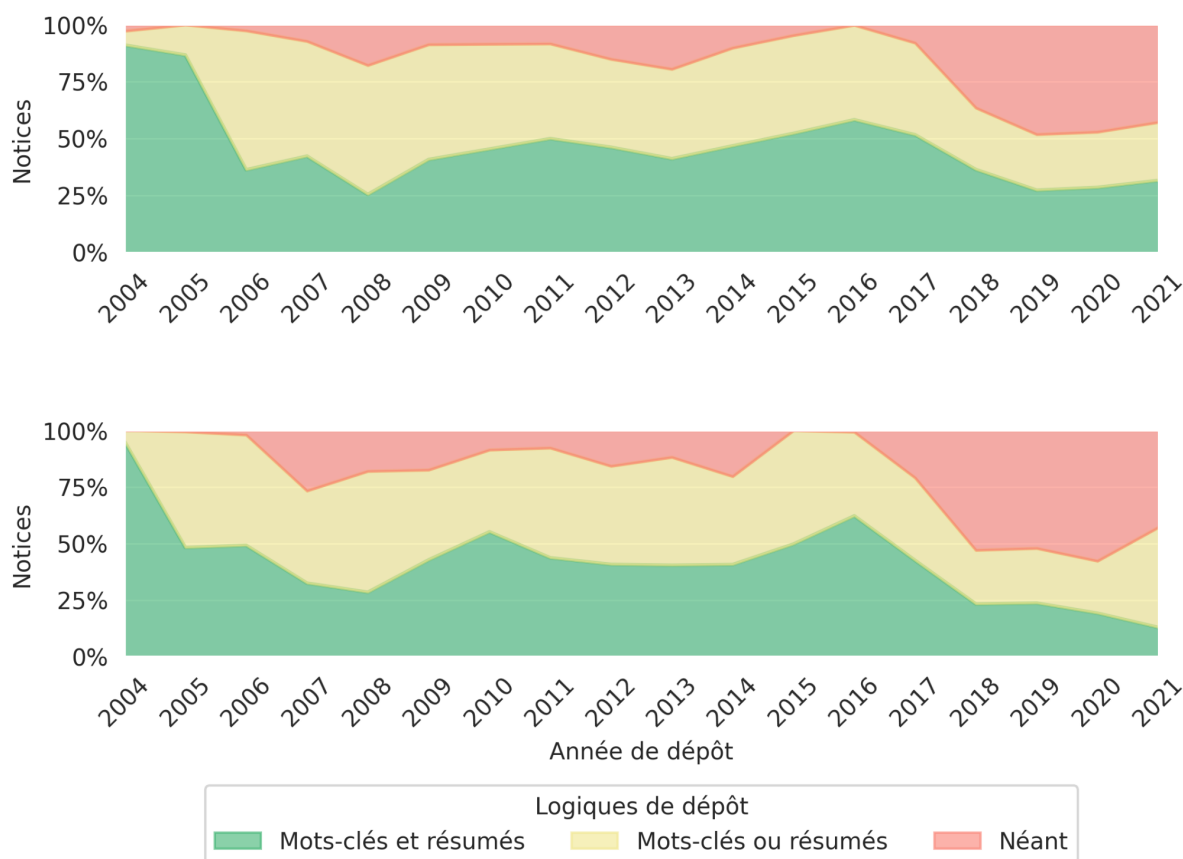
La figure 39 illustre l'évolution des pratiques de description des dépôts en sciences humaines et sociales (SHS), en différenciant les contributeurs en auto-archivage (haut) des contributeurs intermédiaires (bas), sur la période couvrant 2004 à 2021.

Pour les contributeurs réguliers, la proportion de dépôts décrits par à la fois des mots-clés et des résumés représente une large majorité en début de période avec près de 90 % des dépôts en 2004. Cette proportion chute dès 2006 à près de 35 %, les contributeurs préférant les dépôts décrits par des mots-clés ou des résumés qui représentent environ 60 % des dépôts la même année. Jusqu'en 2017, on observe une tendance progressive vers l'amélioration de la description des nouveaux dépôts. En 2017, les dépôts décrits par des résumés et des mots-clés représentent un peu plus de 50 % des dépôts. Néanmoins, nous retrouvons toujours le même décalage qui se produit dès 2018 où la part de dépôts non décrite augmente fortement concomitamment avec l'augmentation du nombre de dépôts. Ainsi, en 2021, environ 32 % des dépôts sont décrits par à la fois des résumés et des mots-clés.

Pour les nouveaux contributeurs, à l'image des figures précédemment décrites, la tendance est quasiment identique avec un moindre degré de description, à l'exception d'une période de fluctuation plus trouble dans les années 2010. Sur la fin de la période observée, la proportion de notices non décrites est d'environ 42%, identiquement que pour les types de contributeurs étudiés ci-dessus, cependant, le niveau de notices complètes, c'est-à-dire, décrites par à la fois des résumés et des mots-clés montre une nette tendance à la baisse sur les dernières années représentant moins de 20% des dépôts en 2021. On constate donc que sur la fin de la période, les nouveaux contributeurs apportent une moindre attention à la description des dépôts.

Figure 39

Évolution des niveaux de description des dépôts dans le temps en SHS déposées par des contributeurs réguliers (haut) et par des nouveaux contributeurs (bas)



3.2.4. Discussion intermédiaire

L'analyse des pratiques d'archivage en sciences humaines et sociales (SHS) met en lumière des dynamiques singulières, révélatrices des spécificités épistémologiques et disciplinaires propres à ce domaine. Contrairement aux sciences, techniques et médecine (STM), où les logiques de diffusion et de conservation des résultats scientifiques dominant, les SHS montrent une diversité de pratiques marquées par une prédominance des logiques de recensement et de référencement. Cette distinction témoigne d'un ancrage disciplinaire davantage tourné vers la documentation des travaux passés et la mise en visibilité institutionnelle plutôt que vers une diffusion immédiate et exhaustive des résultats.

Là encore, les changements de pratiques repérés s'articulent avec des évolutions institutionnelles de HAL, telles que le développement de portails dédiés aux établissements ou la communication accrue autour du dépôt via les services de documentation. La création, en 2005, de l'archive

HAL-SHS dédiée aux SHS constitue un tournant structurant. En offrant un espace spécifiquement pensé pour les disciplines SHS, cette initiative a contribué à favoriser l'appropriation du dépôt par les chercheurs de ces domaines et à légitimer des formes de dépôt adaptées à leurs objets et pratiques scientifiques. Elle a aussi participé à infléchir les logiques d'usage de la plateforme HAL, en renforçant le poids des dépôts à visée de référencement plutôt que de diffusion intégrale.

Les données révèlent trois phases dans l'adoption des pratiques d'archivage, avec une accélération notable à partir de 2017, soutenue par les initiatives en faveur de la science ouverte. Cependant, cette période est aussi caractérisée par une diminution préoccupante des niveaux de description des dépôts, avec une augmentation significative des notices non décrites. Ce phénomène contraste avec les efforts constatés dans les STM et soulève des enjeux pour la découvrabilité et l'accessibilité des travaux en SHS. Les contributeurs, qu'ils soient réguliers ou nouveaux, semblent accorder une moindre priorité à la complétude des métadonnées, ce qui pourrait limiter l'impact scientifique des dépôts.

La diversité des logiques de dépôt, dominée par le recensement et le référencement, reflète les particularités des SHS, où les cycles de publication et les traditions disciplinaires influencent fortement les pratiques. Cependant, cette orientation vers des dépôts sans texte intégral ni métadonnées complètes pourrait poser des défis dans le contexte actuel de la science ouverte, où l'accent est mis sur la transparence et la visibilité des recherches.

En somme, la cartographie des pratiques en SHS révèle une communauté scientifique en transition, c'est-à-dire engagée dans un processus de recomposition progressive de ses pratiques, sous l'effet conjugué des politiques de science ouverte, des injonctions à la visibilité, et de la montée en puissance des infrastructures numériques. Cette transition ne se réduit pas à une opposition entre anciens et nouveaux modèles, mais traduit une forme de tension entre des logiques ancrées dans la culture disciplinaire des SHS, centrée sur le temps long, la monographie, la maîtrise du sens, et les nouvelles attentes en matière de diffusion rapide, d'accessibilité, et de valorisation quantitative des productions. Ces résultats soulignent la nécessité de stratégies adaptées pour soutenir les chercheurs en SHS dans leurs efforts d'archivage, en tenant compte de leurs spécificités tout en encourageant des pratiques qui améliorent la qualité, et par extension, l'impact des dépôts. Cette analyse met également en évidence l'importance de ne pas comparer directement les pratiques des SHS à celles des STM, chaque domaine ayant des dynamiques et des besoins propres.

3.2.5. Des disparités notables dans les pratiques d'archivage entre les disciplines des sciences, techniques et médecine et des sciences humaines et sociales

Les pratiques d'archivage diffèrent sur de nombreux points entre les deux grands ensembles de disciplines que sont les sciences, techniques et médecine (STM) et les sciences humaines et sociales (SHS). Depuis 2017, les dépôts en SHS sont bien moins décrits que ceux relevant des STM. Là où les disciplines STM privilégient de plus en plus des dépôts correctement décrits, incluant des mots-clés et des résumés pour assurer une meilleure visibilité et accessibilité des travaux, les SHS semblent avoir relâché leurs efforts en matière de description. En effet, la proportion de dépôts en SHS sans aucune description a augmenté de manière significative, surpassant parfois celle des dépôts correctement décrits. En SHS, les dépôts sont également effectués plus souvent selon des logiques de recensement et de référencement, deux approches qui, bien qu'elles contribuent à la documentation des recherches, présentent des intérêts limités pour la diffusion de l'information scientifique. Contrairement aux STM, où l'archivage selon des logiques de communication scientifique directe et d'archivage représente une majorité des dépôts, les SHS semblent accorder une moindre priorité à la diffusion des résultats de recherche. Cette tendance pourrait être attribuée à divers facteurs, tels que des différences dans les cycles de publication, la nature des recherches ou encore les traditions disciplinaires. Ce dernier point est particulièrement important, les différences entre ces ensembles de disciplines sont nombreuses et profondes. Bien que ces variations soient perceptibles en parcourant cette thèse, nous avons choisi de ne pas les illustrer graphiquement dans cette section. Cette décision souligne notre conviction que toute comparaison directe entre ces disciplines resterait délicate et potentiellement trompeuse. Il est essentiel de reconnaître que chaque domaine possède ses propres dynamiques et contextes, rendant toute juxtaposition difficile à interpréter de manière juste et équitable.

3.2.6. Les sciences de l'information et de la communication

Nous abordons à présent les sciences de l'information et de la communication, un domaine qui se situe à l'intersection des technologies, des médias et des interactions humaines. Cette section de l'analyse met en exergue les particularités propres aux pratiques d'archivage dans ce champ spécifique au sein des sciences humaines et sociales. Les sciences de l'information et de la communication, avec leurs approches interdisciplinaires, sont caractérisées par des dynamiques de contribution qui reflètent à la fois les avancées technologiques et les contextes sociaux dans lesquels elles s'inscrivent.

Le premier graphique (figure 40) illustre la proportion des dépôts en sciences de l'information et de la communication (SIC) dans les sciences humaines et sociales (SHS) au fil des années. Le second graphique (figure 41) présente la proportion des contributeurs en sciences de l'information et de la communication dans le même ensemble précédemment mentionné.

La proportion des dépôts en SIC dans le corpus parent affiche une nette domination pour les premières années de la période observée. En 2002, environ 28,5 % des dépôts sont en SIC, atteignant un pic à 36,6 % en 2003, avant de retomber significativement à partir de 2004. En 2005, cette proportion connaît une baisse marquée pour atteindre 4,7 % des dépôts de l'ensemble. Par la suite, la proportion reste faible, oscillant entre 1 % et 3 % pour les années suivantes.

La dynamique observée du côté des contributeurs suit une tendance similaire à celle des dépôts. En 2002, 61,5 % des contributeurs sont issus des sciences de l'information et de la communication, un chiffre qui atteint 72,1 % en 2003. Cette part décroît ensuite drastiquement, avec une chute notable dès 2004 où seulement 42,2 % des contributeurs appartiennent à cette discipline. En 2005, la proportion est réduite à 17,5 % et continue de diminuer pour les années suivantes, atteignant des valeurs inférieures à 10 % à partir de 2007.

Ces résultats révèlent un engagement initial particulièrement marqué de la communauté des SIC dans les premières phases de l'archivage en France. Durant cette période, la discipline se positionne comme un moteur central, contribuant de manière significative à l'ensemble des dépôts dans l'ensemble parent. Toutefois, cette dynamique s'essouffle dès 2004, marquant, à la fois, un recul progressif de l'implication des SIC dans le processus d'archivage et un démarrage plus tardif de l'archivage au sein des autres disciplines des SHS.

Figure 40

Évolution de la proportion de dépôts en SIC dans le corpus de données SHS, dans le temps

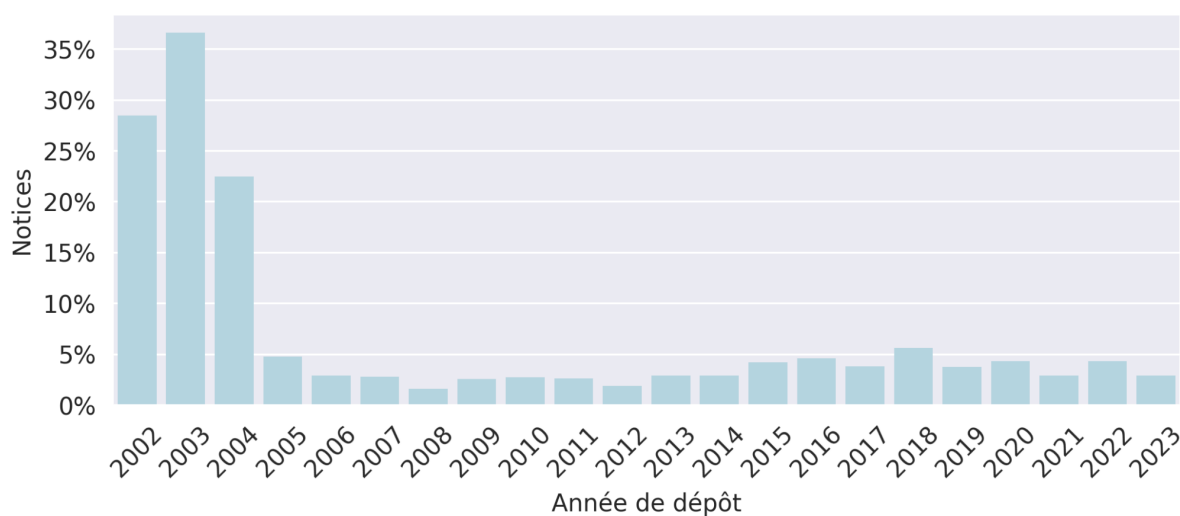
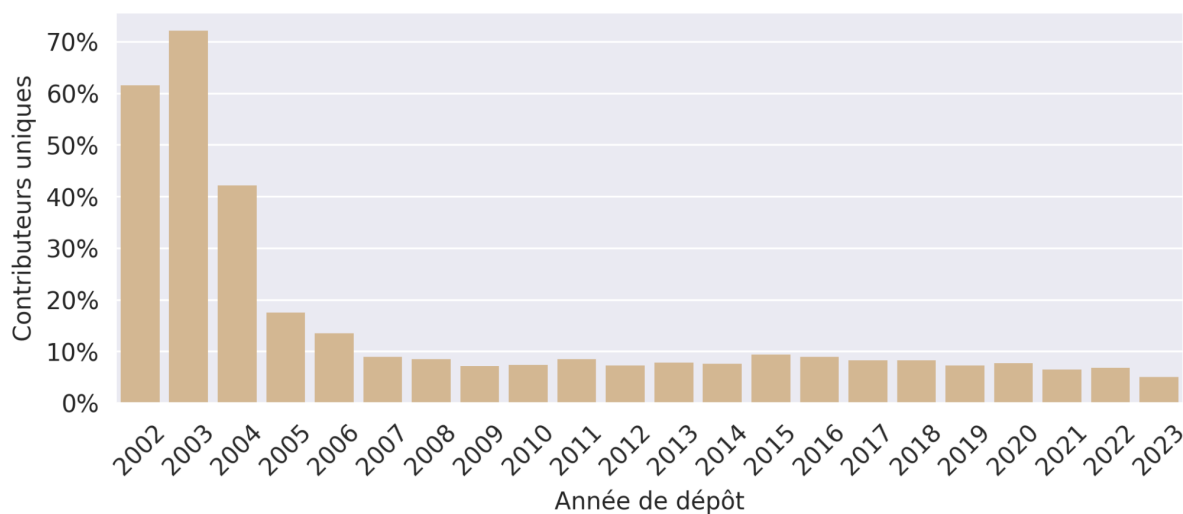


Figure 41

Évolution de la proportion de contributeurs en SIC dans le corpus de données SHS, dans le temps



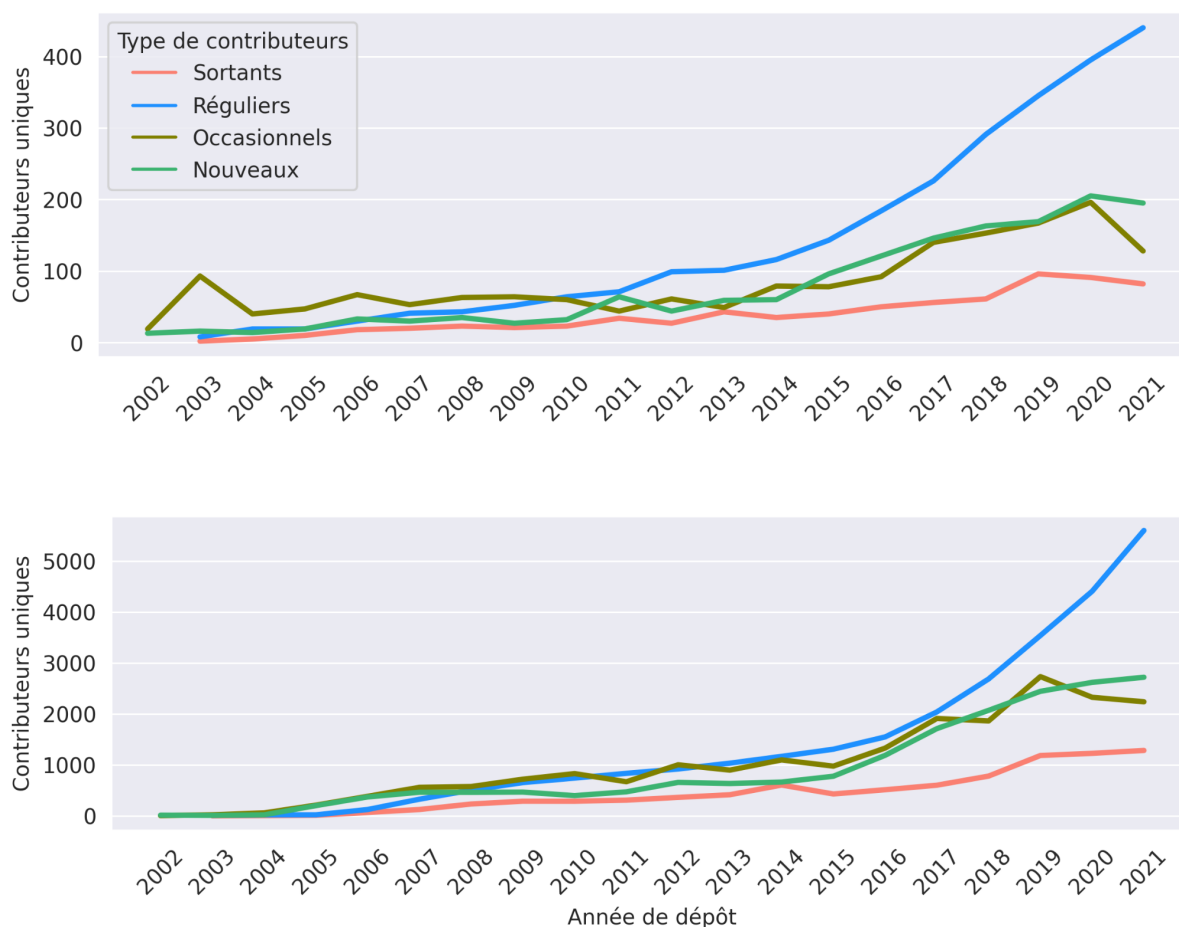
Les deux graphiques de la figure 42 illustrent l'évolution des types de contributeurs dans le temps, comparant les sciences de l'information et de la communication (SIC) en haut, et les sciences humaines et sociales (SHS), SIC exclues en bas. Ces graphiques comparent quatre types de contributeurs : les réguliers (ligne bleue), les occasionnels (ligne verte), les nouveaux (ligne jaune), et les sortants (ligne rouge).

Dans les SIC, le début de la période est largement dominé par les contributeurs occasionnels, marquant une phase d'expérimentation et de faible engagement régulier dans les pratiques d'auto-archivage. Cette phase est probablement liée à la dynamique initiée par @rchiveSIC dans la discipline. Cependant, à partir de 2010, cette tendance s'inverse, le nombre de contributeurs réguliers commence à croître de manière plus soutenue pour devenir le premier type de contributeurs. Dès lors, la courbe des réguliers ne cesse d'évoluer de manière exponentielle, signe d'une adhésion croissante à des pratiques d'auto-archivage récurrentes. En parallèle, l'acquisition de nouveaux utilisateurs reste constante, contribuant à alimenter cette dynamique de croissance, même si nous constatons un possible début de repli en 2021.

En comparaison, en SHS, nous ne retrouvons pas cette phase d'expérimentation au début des années 2000, ce qui conforte l'idée d'une dynamique portée manifestement par le projet @rchiveSIC au sein des SIC. On observe également un décalage plus tardif en SHS entre les contributeurs réguliers et occasionnels. Jusqu'en 2017, les deux courbes suivent une trajectoire similaire en SHS, tandis que dans les SIC, le nombre de contributeurs réguliers se distingue clairement des autres types de contributeurs dès 2011. Cette différence met en lumière une adoption plus précoce des pratiques d'archivage régulier dans les SIC, en contraste avec une évolution plus tardive dans les SHS.

Figure 42

Évolution des types de contributeurs dans le temps en SIC (haut) et en SHS hors SIC (bas)



La figure 43 illustre l'évolution des logiques de dépôt dans le temps pour les sciences de l'information et de la communication (SIC) et les sciences humaines et sociales hors SIC (SHS), en comparant quatre logiques de dépôt.

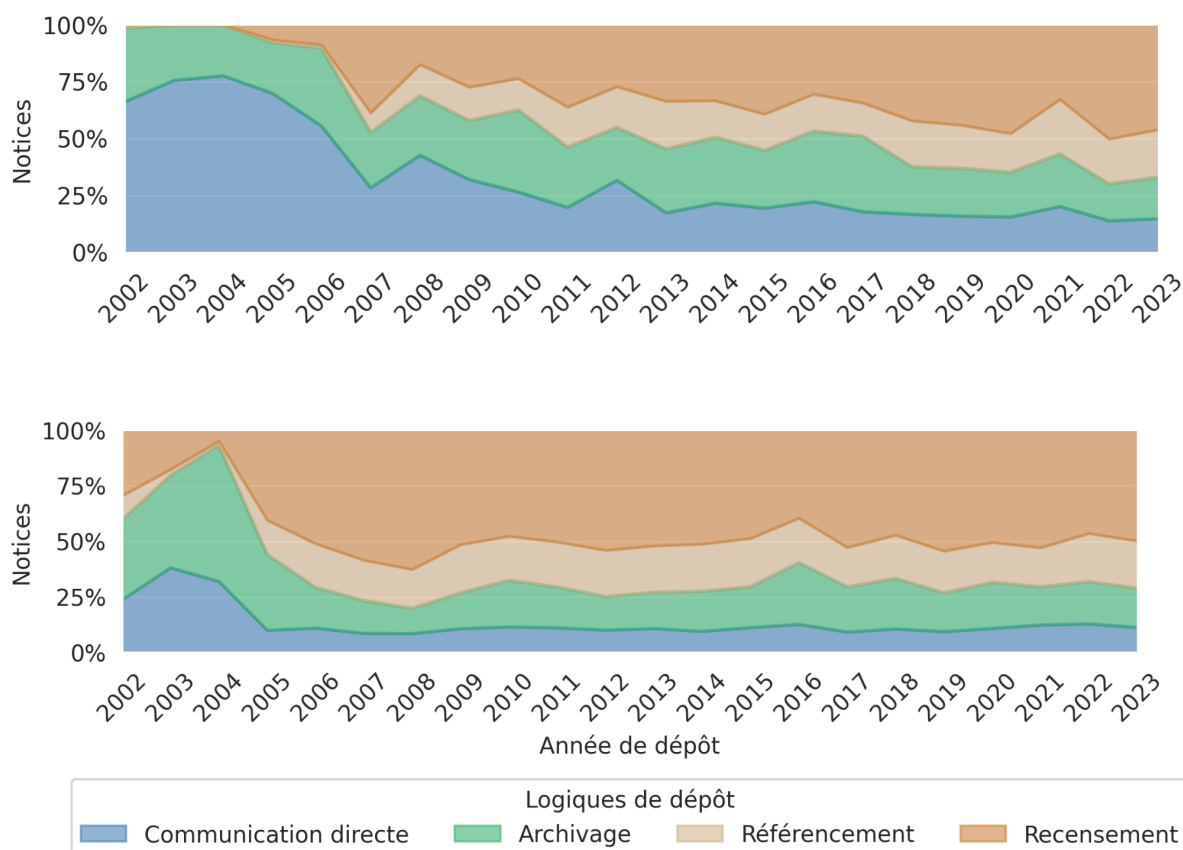
Dans les SIC, la période initiale est largement dominée par la communication scientifique directe, représentant près de 66 % des dépôts en 2002. Cette tendance se poursuit durant les premières années avec des pics encore plus importants en 2003 et 2004, où cette logique représente jusqu'à 77 % des dépôts. Toutefois, à partir de 2006, on observe une répartition plus équilibrée des logiques de dépôt, marquant une diversification des pratiques. À partir de cette année-là, et à l'exception de 2007, les parts de dépôts effectués dans des logiques de recensement et de référencement ne cessent de croître, remplaçant progressivement la part des dépôts effectués dans une logique de communication scientifique directe et, dans une moindre mesure, la part des logiques effectués dans une logique d'archivage. En fin de période, l'archivage représente environ 18,3 % des dépôts, tandis que la communication scientifique directe conserve une part de 14,8 %

des dépôts. Le recensement domine avec près de 46 % des dépôts et le référencement atteint 20,8 % des dépôts. La proportion prépondérante des deux dernières catégories confirme une tendance à déposer dans des logiques de simple visibilité des travaux sans document associé, à l'inverse des observations constatées en début de période.

En SHS, la dynamique diffère particulièrement en début de période. En 2002, la communication scientifique directe représente une part importante (23,7 %), mais c'est l'archivage qui domine avec 36 %. La part des dépôts de recensement est déjà significative à 29,6 %, un chiffre beaucoup plus élevé que dans les SIC à la même époque. Contrairement aux SIC, où la communication directe reste largement prépondérante au début, les SHS montrent dès les premières années une plus grande diversité dans les logiques de dépôt. Dès 2005, on observe une progression plus nette de la part des dépôts effectués selon une logique de recensement qu'en SIC. En fin de période, en 2023 pour les SHS, la répartition des logiques de dépôt présente une situation similaire à celle des SIC, avec une prédominance du recensement qui représente environ 49,9 % des dépôts. L'archivage, bien que moins dominant, occupe tout de même 17,6 % des dépôts, marquant une reprise modérée après avoir été en déclin au milieu de la période. La communication scientifique directe reste relativement stable avec une part de 11,2 %, tandis que le référencement atteint 21,2 %. La différence observée en début de période montre que la dynamique initiée au sein des SIC s'est rapidement orientée sur des logiques d'archivage et particulièrement de communication scientifique directe, pertinentes pour la diffusion des travaux de recherche. La différence dans le rythme de l'évolution vers des répartitions plus stables en fin de période semble souligner que la dynamique initiée auparavant semble impacter positivement le processus de dépôt sur plusieurs années.

Figure 43

Évolution des logiques de dépôt dans le temps en SIC (haut) et en SHS hors SIC (bas)



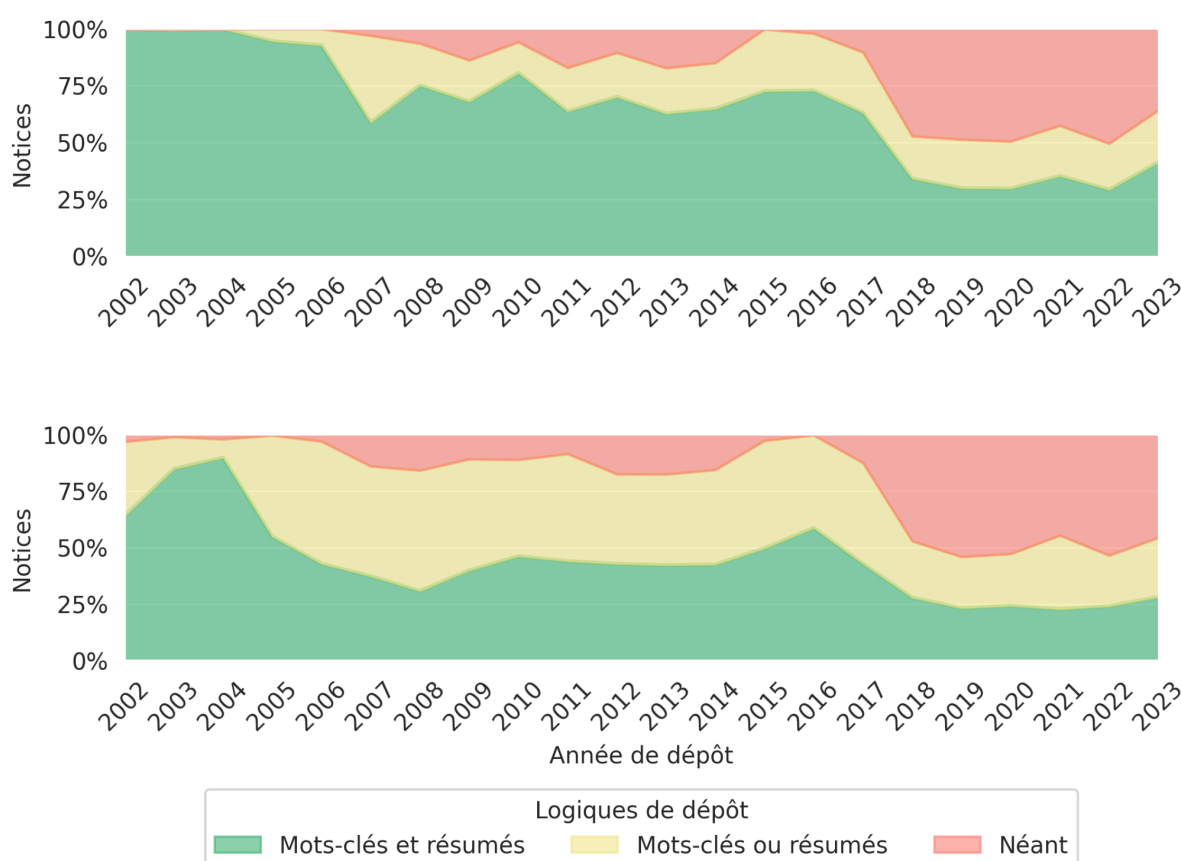
La figure 44 présente l'évolution des niveaux de description des dépôts dans le temps pour les sciences de l'information et de la communication (SIC) et les sciences humaines et sociales (SHS SIC exclues), répartis selon trois niveaux de description, à savoir, les dépôts décrits par des mots-clés et des résumés (vert), des mots-clés ou des résumés (jaune), ou non décrits (rouge).

Dans les SIC, la période initiale (2002 à 2005) montre une très forte dominance des dépôts avec à la fois mots-clés et résumés, représentant presque l'intégralité des notices déposées. Cependant, à partir de 2006, un changement progressif s'opère. On observe une augmentation progressive des dépôts non décrits ou ne comportant que mots-clés ou des résumés. Ces proportions restent marginales jusqu'en 2018 où on observe une chute brutale dans le niveau de description global des dépôts. En 2018, la part des notices décrites par à la fois des mots-clés et des résumés ne représente plus que 34,3 % des dépôts. Les proportions restent plutôt stables dans les années qui suivent, et on observe une légère amélioration en 2023, où cette catégorie atteint 41,5 %. Cela suggère une timide reprise progressive des bonnes pratiques de description des dépôts après une période de dégradation notable.

Dans les SHS, la dynamique est différente. En 2002, près de 64 % des dépôts incluent à la fois des mots-clés et des résumés, tandis qu'environ 32 % n'en contiennent que l'un des deux. Cette première proportion augmente jusqu'en 2004, atteignant 89,8 % des dépôts, avant de connaître une baisse significative pour tomber à 30,8 % des dépôts en 2008. À l'inverse des SIC, la proportion de dépôts correctement décrits progresse ensuite jusqu'en 2016 avec 58,7 % des dépôts avant de connaître une chute dans le niveau de description global similaire à celle observée dans les SIC dès 2018. Là encore, l'excellent niveau de description observé dans les SIC jusqu'en 2018 illustre la bonne dynamique initiée au sein de la discipline dès les premières années d'observation.

Figure 44

Évolution des niveaux de description des dépôts dans le temps en SIC (haut) et en SHS hors SIC (bas)



3.2.7. Discussion intermédiaire

L'analyse des pratiques d'archivage dans les sciences de l'information et de la communication (SIC) révèle des dynamiques distinctes et une influence déterminante dans les premières années

de l'adoption de l'archivage en France. Ces pratiques, caractérisées par une prédominance initiale de la communication scientifique directe, témoignent d'une volonté forte de diffusion immédiate des résultats, soutenue notamment par l'initiative @rchiveSIC. Cette dynamique a positionné les SIC comme un moteur majeur du développement des pratiques d'archivage dans les sciences humaines et sociales (SHS), en particulier au début des années 2000.

Toutefois, la période post-2006 marque une transition notable. 2006 est également l'année à partir de laquelle @rchiveSIC est hébergée sur HAL, ce qui a pu contribuer à infléchir les dynamiques d'archivage au sein de la discipline. Une diversification progressive des logiques de dépôt s'installe, avec une montée en puissance des logiques de recensement et de référencement au détriment de la communication scientifique directe. Parallèlement, les niveaux de description des dépôts, initialement exemplaires, connaissent une dégradation significative après 2018, reflétant une évolution préoccupante vers des pratiques moins rigoureuses. Cette tendance, bien que alignée avec des dynamiques similaires dans d'autres disciplines des SHS, souligne une perte partielle de la priorité accordée à la complétude des métadonnées, compromettant ainsi la visibilité et l'accessibilité des travaux.

Les SIC se distinguent néanmoins par une adoption précoce et structurée des pratiques d'archivage régulier. Les contributeurs réguliers prennent rapidement le relais des contributeurs occasionnels, contrastant avec une évolution plus tardive dans d'autres disciplines des SHS. Cette transition témoigne d'une appropriation des outils d'archivage favorisée par les spécificités interdisciplinaires des SIC, où les interactions entre technologies, médias et humanités numériques jouent un rôle central.

En somme, les SIC illustrent un cas exemplaire d'engagement initial dans les pratiques d'archivage, suivi d'une phase de diversification et de consolidation. Cependant, les récents défis liés à la qualité des descriptions de dépôts appellent à un renforcement des stratégies de sensibilisation et de formation, pour préserver l'héritage de cette discipline pionnière et maximiser l'impact de ses contributions dans l'écosystème des archives ouvertes. Ces constats soulignent également l'importance de tirer parti des enseignements des spécificités des SIC au début de la période observée pour inspirer de nouvelles initiatives similaires dans d'autres disciplines des SHS.

3.2.8. *L'Université de Toulon*

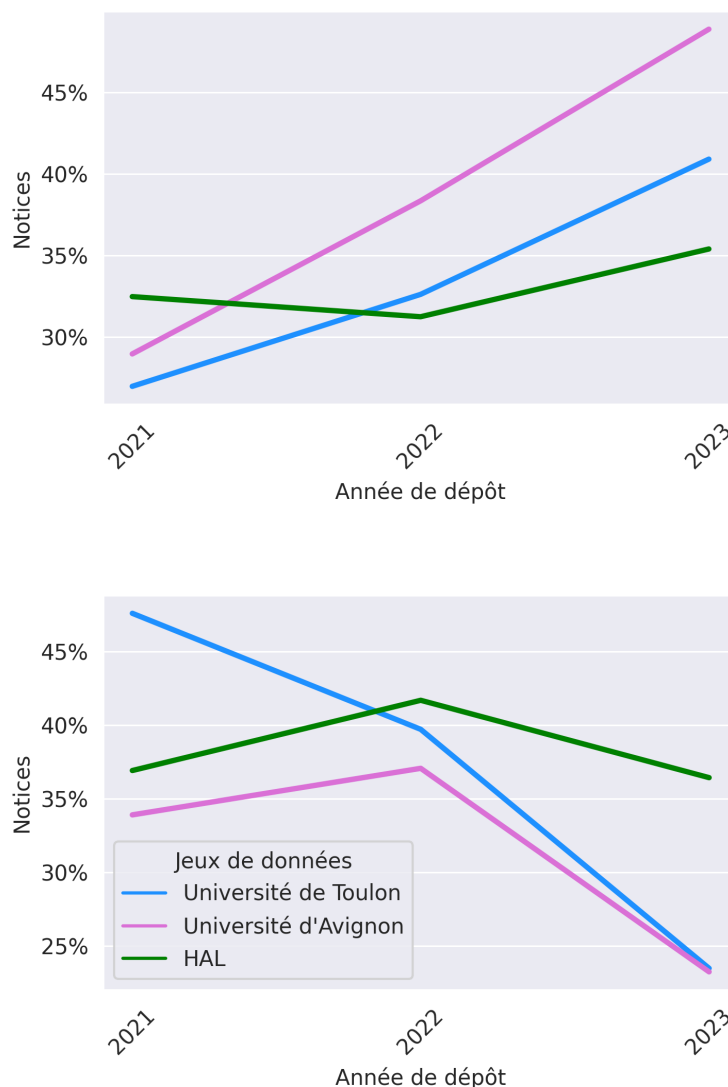
Nous concluons notre exploration par une étude focalisée sur les notices tamponnées par l'Université de Toulon. Cette analyse porte sur l'impact des politiques locales, notamment le

dispositif SoVisu, sur les pratiques de dépôt des chercheurs de l'université. Ainsi, dans cette partie, nous mettons en avant les différences entre tendances au niveau de l'intégralité de l'archive HAL et au sein des universités de taille similaire. Les stagiaires de l'École nationale supérieure des sciences de l'information et des bibliothèques (Enssib) rencontrés lors d'une réunion courant 2023 à l'Université de Toulon, nous indiquent une liste des établissements français similaires à l'Université de Toulon, à savoir, les universités de Pau et des Pays de l'Adour, Savoie Mont-Blanc ou encore celle d'Avignon (Cacheux et al., 2023). En termes de présence sur HAL, l'Université d'Avignon se présente comme l'établissement le plus comparable à Toulon, notamment en raison de leur similitude en volume de dépôts et de leur répartition disciplinaire. Nous basons donc les comparaisons ci-après sur cette université. En mettant en lumière les effets positifs de ces initiatives, nous examinons comment les politiques et projets spécifiques à l'Université de Toulon ont amélioré les processus d'archivage, par extension de découvrabilité des travaux et donc de diffusion de l'information scientifique.

En observant la figure 45, nous constatons une nette amélioration du niveau de description des notices à partir de l'année 2022, correspondant à la date de lancement du projet. Pour les notices tamponnées par l'Université de Toulon, la part de notices décrites par à la fois des mots-clés et des résumés progresse nettement, évoluant d'environ 27 % en 2021 à environ 41 % en 2023. Cette progression est légèrement moins marquée que pour les notices tamponnées par l'Université d'Avignon, qui a enregistré une progression de 20 points sur la même période. En revanche, cette dynamique contraste avec la tendance globale de l'archive HAL, où le même niveau de description des notices stagne aux alentours de 33 %. Les évolutions des notices non décrites complètent ces observations. La proportion de notices non décrites tamponnées par l'Université de Toulon chute fortement, passant d'environ 48 % en 2021 à environ 23 % en 2023. Ce recul est nettement plus marqué que pour les notices tamponnées par l'Université d'Avignon où il n'est que de 10 points sur la même période. Les comparaisons avec l'Université d'Avignon sont toutefois limitées par le manque d'informations sur les initiatives et politiques locales entreprises. De plus, comme démontré précédemment, les pratiques diffèrent considérablement selon les disciplines étudiées, et, bien que les répartitions disciplinaires des deux établissements s'approchent, elles ne sont pas identiques, introduisant ainsi un bruit dans la comparaison. Nous ne développons pas davantage l'analyse de ces éléments, mais nous avons jugé pertinent de présenter ces informations afin de fournir un contexte et une compréhension plus complète des dynamiques en jeu.

Figure 45

Évolution des notices décrites par à la fois un résumé et des mots-clés (haut) et celles non décrites (bas) depuis 2021



Ces résultats suggèrent que le déploiement de la plateforme SoVisu, dans le cadre des dynamiques des politiques locales mises en œuvre, a un impact positif sur les pratiques des chercheurs et par extension sur l'appropriation de la science ouverte par les chercheurs de l'Université de Toulon. L'augmentation du nombre de notices richement décrites, accompagnée de la baisse inéluctable des notices peu descriptives, indique d'une amélioration des pratiques de métadonnées. Cette tendance, qui contraste nettement avec la tendance générale de l'archive, démontre l'efficacité des initiatives locales entreprises, dans le contexte décrit précédemment, pour favoriser la diffusion des travaux toulonnais. Ce graphique constitue, en ce sens, un instrument clé de mesure d'impact. Il permet d'objectiver l'effet d'un accompagnement structuré

sur la qualité des dépôts, à travers des indicateurs concrets tels que la part de notices enrichies ou le recul des notices non décrites. Une déclinaison de cette analyse à l'échelle des laboratoires, voire des disciplines, constituerait un prolongement méthodologique pertinent. Elle permettrait de mieux évaluer la portée réelle des dispositifs de soutien, en croisant les dynamiques d'accompagnement avec des indicateurs d'usage et de qualité documentaire. Bien qu'une enquête qualitative, par le biais d'entretiens, permettrait uniquement d'identifier et d'analyser les impacts distincts de l'ensemble des initiatives, il n'est pas forcément souhaitable de chercher à isoler les effets de chacune d'entre elles. Nous estimons en effet que leur utilisation conjointe est essentielle pour maximiser leur utilité. Dans ce contexte, les interactions entre ces différentes politiques et projets peuvent générer des synergies positives, rendant leur impact combiné plus significatif que la somme de leurs effets individuels.

3.3. Discussion

Les résultats présentés dans ce travail offrent un éclairage complémentaire aux indicateurs institutionnels à portée nationale, tels que le Baromètre de la science ouverte (Jeangirard, 2019), en proposant une lecture fine des dynamiques d'archivage au sein de l'archive HAL. Là où ces baromètres s'attachent principalement à mesurer la proportion de publications disponibles en libre accès, toutes plateformes confondues, notre approche vise à qualifier les pratiques effectives de dépôt et à analyser les logiques d'action des contributeurs. Elle permet ainsi d'interroger non seulement le degré d'ouverture, mais aussi les modalités concrètes de cette ouverture, à travers la complétude des métadonnées, les types de documents déposés et les usages professionnels de la plateforme.

Cette analyse confirme que si HAL n'est pas présentée comme une simple base de données, son fonctionnement dépasse largement cette catégorie. Elle opère comme un dispositif info-communicationnel, au sens proposé par Couzinet (2009), c'est-à-dire un espace d'interactions où s'articulent des dimensions techniques (interface, métadonnées, interopérabilité), sociales (acteurs, usages, régulations) et symboliques (valeurs attribuées à la visibilité, à l'ouverture, ou encore à la reconnaissance scientifique). En cela, HAL constitue un objet d'étude pertinent pour les SIC, dans sa capacité à médiatiser des formes de communication scientifique et à inscrire des pratiques individuelles dans un cadre socio-technique structurant.

Les disparités observées entre disciplines peuvent être interprétées, non comme des retards ou des résistances, mais comme des rapports différenciés à l'écriture scientifique et à sa diffusion

(2007). Ce sont ainsi des représentations du savoir, des normes de légitimité et des régimes d'énonciation qui se rejouent dans les usages différenciés de HAL. À ce titre, la sous-représentation des dépôts complets dans certaines disciplines peut refléter un attachement à des formes éditoriales spécifiques (chapitres, ouvrages) ou à des modèles de diffusion plus traditionnels, difficilement compatibles avec les standards de l'archive.

Par ailleurs, les types de contributeurs identifiés, réguliers, entrants, sortants, renvoient à des profils d'appropriation différenciés, eux-mêmes influencés par des facteurs institutionnels (politiques d'établissement, rôle des gestionnaires de collection) ou personnels (compétence informationnelle, rapport à la visibilité). Dans cette perspective, les logiques d'auto-archivage apparaissent comme autant de prises de position face aux transformations contemporaines du système de publication scientifique, et notamment face à la montée en puissance des dispositifs d'évaluation numérique (Broudoux, 2017). Ainsi, HAL n'est pas un simple outil d'archivage : elle est aussi un espace de mise en visibilité de soi comme chercheur et un instrument possible de la fabrique identitaire académique. La manière dont les chercheurs s'approprient, ou non, cette plateforme peut dès lors être lue comme une stratégie, parfois implicite, de négociation avec les nouvelles normes de la science ouverte. Dans cette perspective, il pourrait être pertinent d'approfondir la typologie des contributeurs. À l'instar des travaux de Mahé et Prime-Claverie (2017a) qui distinguent les chercheurs des personnels d'appui à la recherche, tels que les bibliothécaires, documentalistes ou gestionnaires de collections, la déclinaison des contributeurs intermédiaires permettrait d'affiner la compréhension des usages différenciés de HAL. Appliquer une telle différenciation à l'échelle des corpus étudiés ouvrirait la voie à une lecture plus nuancée des pratiques d'appropriation, en articulant logiques professionnelles, contextes institutionnels et compétences informationnelles.

En creux, cette étude met également en lumière certaines limites : la qualité des métadonnées reste inégale, et les pratiques de dépôt, bien qu'en croissance, peinent encore à se stabiliser durablement. Les constats exposés ici rejoignent les conclusions d'auteurs de la discipline. Schöpfel et al. (2022) indiquent notamment que « les chercheurs utilisent HAL, peut-être moins pour chercher de l'information mais surtout pour signaler et diffuser leurs travaux, et les laboratoires jouent un rôle important pour la mise en œuvre de cette politique ». Ces constats posent la question de l'accompagnement des chercheurs, non seulement sur le plan technique, mais aussi sur le plan culturel et informationnel. Il ne suffit pas d'ouvrir un espace pour que celui-ci soit occupé, encore faut-il en outiller l'usage, comme le montre le développement du dispositif SoVisu et l'analyse à posteriori de ses effets sur les pratiques d'archivage. Ces dispositifs

peuvent être compris comme des « dispositifs de gouvernementalité documentaire », au sens où ils prescrivent des formes de visibilité, structurent des hiérarchies d'accès, et configurent implicitement des normes de conformité (Chartron, 2010). En cela, les métadonnées ne sont pas de simples descripteurs mais des agents informationnels, qui participent à la fabrique d'une science archivable, interopérable, et donc gouvernable.

Lorsqu'on met en parallèle l'évolution des pratiques de dépôt avec la frise chronologique des grandes étapes du développement de HAL, on constate que ces pratiques sont régulièrement influencées par les choix structurants de l'archive. L'apparition de nouvelles fonctionnalités, la simplification de l'interface de dépôt, la création de portails disciplinaires ou encore l'intégration de référentiels ont un impact direct sur les contributeurs et leurs façons d'utiliser la plateforme. Autrement dit, les décisions de conception technique et d'orientation stratégique prises par le CCSD modèlent les usages, en favorisant certaines pratiques au détriment d'autres. Cette articulation entre infrastructure et pratiques rappelle que l'appropriation d'un dispositif ne saurait être dissociée de sa configuration technique et organisationnelle, et que la gouvernance de l'archive joue un rôle central dans l'orientation des usages et la fabrique des comportements scientifiques. Cette évolution des fonctionnalités et de l'ergonomie, en facilitant l'accès au dépôt, a également contribué à élargir le spectre des utilisateurs. Si cela constitue une avancée en termes d'ouverture et d'inclusivité, cela peut aussi expliquer la persistance de pratiques de dépôt incomplètes ou hétérogènes. Certains contributeurs, nouvellement engagés dans l'auto-archivage, ne perçoivent pas encore pleinement les enjeux liés à la qualité du dépôt, notamment en termes de repérabilité de leurs travaux dans les moteurs de recherche, de valorisation de leur production scientifique ou de visibilité académique. D'où l'importance de dispositifs d'accompagnement adaptés, non seulement techniques, mais aussi informationnels, pour soutenir une appropriation de qualité du dépôt en libre accès.

Toutefois, les évolutions observées à partir des traces numériques ne suffisent pas à restituer la complexité des usages. Si l'on peut identifier des profils de contributeurs et des tendances d'appropriation, les raisons pour lesquelles certains choisissent de déposer (ou non), et la manière dont ils perçoivent les bénéfices attendus du dépôt, demeurent largement implicites. C'est pourquoi il importe d'interroger les dimensions subjectives et situées de ces pratiques, au-delà des seules données quantitatives. Malgré la richesse des données mobilisées et la solidité des corpus analysés, la profondeur interprétative des résultats appelle cependant quelques réserves. Si les tendances statistiques permettent d'identifier des logiques de dépôt et de typologies de contributeurs, le sens que les acteurs attribuent à leurs propres pratiques demeure peu explicité.

Cette limite tient à la nature même des traces numériques étudiées, qui ne disent rien des motivations, des contraintes ou des représentations individuelles. En ce sens, l'analyse gagnerait à être prolongée par des approches qualitatives complémentaires, telles que des entretiens semi-directifs ou des analyses de discours, afin de mieux cerner les logiques d'action situées, notamment face aux injonctions contemporaines à la visibilité et à la performance. De plus, les disparités disciplinaires identifiées ne sauraient être réduites à de simples écarts d'appropriation technique : elles renvoient à des cultures éditoriales distinctes, à des régimes de légitimité hétérogènes, et à des contraintes institutionnelles variables. Une lecture plus approfondie, ancrée dans une analyse socio-discursive des disciplines, permettrait d'enrichir cette cartographie. Enfin, il convient de rappeler que les dépôts observés sur l'archive HAL ne représentent qu'une partie des pratiques effectives de diffusion scientifique, et qu'ils peuvent refléter autant une adhésion volontaire à la science ouverte qu'une réponse contrainte à des logiques d'évaluation.

Conclusions et perspectives

« En somme, la découvrabilité n'est pas seulement un enjeu technique, mais une question de sens : celle de la manière dont la recherche se donne à voir, se partage et s'inscrit dans le monde. »

Avant d'approfondir la conclusion de cette thèse, nous souhaitons inviter le lecteur, à la faveur d'une réflexion suscitée par l'écoute d'un audio (Wakim, 2025), à envisager les changements de paradigme auxquels notre société est aujourd'hui confrontée. Qu'il s'agisse, de manière très concrète, du domaine agricole avec la controverse autour d'un éventuel retour des insecticides néonicotinoïdes, ou, comme évoqué précédemment, du monde scientifique en mutation, la problématique demeure fondamentalement la même : aucun changement de paradigme ne peut s'opérer sans heurts, à moins qu'un accompagnement réfléchi, structuré et présent ne soit mis en place. C'est précisément l'objet que ce manuscrit s'attache à explorer.

À partir d'une expérimentation conduite dans le cadre d'une recherche-action, centrée sur la conception et le déploiement du dispositif SoVisu, émergent plusieurs questionnements sur les conditions concrètes d'appropriation de l'archive ouverte HAL. Loin de se réduire à une infrastructure technique, HAL fonctionne comme un espace de négociation entre injonctions institutionnelles, logiques locales d'accompagnement et pratiques individuelles. L'expérimentation montre à la fois le potentiel mobilisateur d'un dispositif de visualisation pensé comme appui à la décision, et les résistances qui peuvent émerger face à ces dispositifs, notamment en contexte de contrainte ou de méconnaissance.

Dans ce contexte, SoVisu se présente comme un outil d'accompagnement à la science ouverte, conçu pour favoriser une prise de conscience des logiques de visibilité à l'œuvre dans les dépôts. Il s'inscrit à ce titre dans une trajectoire plus large d'outillage de l'infrastructure HAL, qui voit apparaître depuis 2020 une série de dispositifs orientés vers l'accompagnement automatisé et la mesurabilité des usages. Le lancement récent de HAL Monitor (CCSD, 2025), qui met à disposition des tableaux de bord d'activité, qui propose des suggestions automatiques de dépôts, illustre cette tendance. Dans cette perspective, SoVisu ne constitue pas une rupture mais une brique complémentaire : il s'adosse à HAL tout en proposant une sur-couche de visualisation et d'analyse critique, pensée pour favoriser l'appropriation humaine des métadonnées. À ce titre, le dispositif aurait vocation, idéalement, à être intégré nativement à l'archive HAL, de manière à renforcer la continuité d'usage et à favoriser son adoption. Son architecture repose néanmoins sur une logique modulaire, ce qui rend possible, à terme, son adaptation à d'autres sources ou entrepôts d'archives ouvertes (comme arXiv ou Zenodo), dès lors que les métadonnées sont exposées selon des standards interopérables. Ce positionnement intermédiaire, entre complémentarité avec HAL et potentiel de portabilité, confère à SoVisu une certaine souplesse, mais pose également la question de son avenir : simple outil connexe ou composant plus structurel des environnements numériques de la science ouverte ? Dans tous les cas, il contribue à

renforcer l'appropriation humaine et critique de données désormais mobilisées à des fins de pilotage et de valorisation institutionnelle. En tant que dispositif de visualisation de l'information, il propose une image miroir des données réellement moissonnées par les systèmes automatisés, dont font partie les moteurs de recherche scientifiques, plateformes d'indexation et agrégateurs. Il donne ainsi à voir, de manière synthétique, ce que les métadonnées révèlent, ou non, des travaux déposés, et permet aux acteurs de mieux saisir les conséquences de leurs choix de saisie sur la circulation des savoirs. Pensé comme un levier de réflexivité, SoVisu articule une dimension pédagogique et stratégique, tout en rendant perceptibles les zones d'ombre ou les points de blocage. Dans ce prolongement, le dispositif SoVisu peut être interprété non seulement comme un outil de visualisation, mais aussi comme un acteur intermédiaire dans la transformation des pratiques d'archivage. En rendant visibles les structures informationnelles des dépôts, il agit comme une image miroir des métadonnées, permettant aux chercheurs, aux services de soutien et aux institutions de prendre conscience de ce qui est réellement exposé, moissonné, ou parfois masqué. Il contribue ainsi à renforcer une littératie des métadonnées, condition indispensable à une appropriation critique des infrastructures de la science ouverte. Ce rôle s'inscrit dans une approche en sciences de l'information et de la communication attentive aux dispositifs de médiation, qui conçoivent la technique non comme un simple outil neutre, mais comme un opérateur de transformation des rapports à la connaissance, aux normes et à la visibilité (Couzinet, 2009 ; Jeanneret & Souchier, 2005). L'expérimentation montre à la fois le potentiel mobilisateur de ce type de dispositif et les résistances qu'il peut susciter, notamment en contexte de contrainte, de surcharge ou de faible acculturation à la logique des métadonnées.

En étudiant les métadonnées associées aux dépôts, et, par extension, en qualifiant l'acte de dépôt et le profil du contributeur, ce travail met en évidence les disparités disciplinaires, les logiques différenciées de contribution, ainsi que les tensions entre prescriptions institutionnelles et usages réels. Les métadonnées descriptives et le texte intégral s'imposent comme un point d'ancrage central : ils conditionnent la visibilité, la traçabilité et la réutilisation des productions scientifiques. Ils constituent aussi un indicateur de l'investissement des chercheurs dans la mise en circulation de leurs travaux, mais également des conditions dans lesquelles ils exercent leur activité. L'analyse des pratiques de dépôt menée dans la troisième partie montre notamment que, malgré une apparente homogénéité des interfaces de dépôt, les pratiques sont profondément contrastées selon les disciplines, les types de documents et les statuts des contributeurs. On observe une hétérogénéité marquée dans la complétude des métadonnées, une variabilité dans la présence du texte intégral, ainsi que des usages parfois éloignés des recommandations institutionnelles. Ces écarts ne traduisent pas nécessairement une négligence, mais révèlent des logiques d'ajustement,

de contournement ou de priorisation, qui s'inscrivent dans des contextes professionnels et informationnels spécifiques. En ce sens, les métadonnées apparaissent comme des traces concrètes d'un arbitrage permanent entre contraintes et volontés de rendre visible le travail scientifique. Elles donnent à voir non seulement la diversité des rapports à l'archive ouverte, mais aussi les formes de participation, plus ou moins actives, à l'économie de la science ouverte.

Cette recherche comporte des limites. Elle ne permet pas d'accéder aux dimensions subjectives des pratiques d'archivage, ni d'observer, isolément, les effets à long terme d'un accompagnement structuré. Ces limites sont d'autant plus marquées que les outils institutionnels documentent peu les retours utilisateurs. C'est précisément dans cet angle mort que s'inscrit la présente thèse, en apportant un éclairage qualitatif sur les pratiques effectives d'appropriation, elle complète les données institutionnelles par une lecture située des usages et des perceptions. Le contexte institutionnel dans lequel elle s'inscrit circonscrit également la portée des résultats. Ces limites ouvrent toutefois des perspectives : prolonger l'analyse par des approches qualitatives, comparer les dynamiques à d'autres environnements disciplinaires ou internationaux, et intégrer l'évolution des outils numériques, notamment ceux basés sur l'intelligence artificielle. Ces transformations interrogent en profondeur les futurs modes de production, de description et de diffusion des savoirs. D'ailleurs, dans ce contexte, l'absence d'un auteur dans les circuits du libre accès ne signifie pas son invisibilité totale, mais une visibilité médiée, potentiellement déformée. En effet, les systèmes d'intelligence artificielle, en s'entraînant majoritairement sur de l'information accessible, construisent une image des chercheurs absents de ces circuits non pas à partir de leurs propres écrits, mais à partir des citations ou commentaires qu'en font d'autres. Cela pose un risque accru de représentation biaisée, où les discours secondaires prennent le pas sur la source première. Ainsi, refuser ou négliger l'archivage revient non seulement à se rendre moins visible, mais aussi à céder le contrôle de la manière dont ses idées seront perçues, indexées et, potentiellement, interprétées par des systèmes automatisés.

Toutefois, attention à ne pas penser l'usage des technologies d'intelligence artificielle comme solution miracle : sans l'information que détient, parfois seul, le chercheur, aucune automatisation ne peut garantir la qualité, la pertinence ou la cohérence des métadonnées produites. L'accompagnement humain, la médiation documentaire et la conscientisation des enjeux restent donc des leviers indispensables pour assurer une ouverture des savoirs à la fois efficace et éthique. En ce sens, cette recherche défend l'idée qu'un accompagnement à la science ouverte ne peut se réduire à des dispositifs techniques, aussi performants soient-ils. Il suppose un travail d'acculturation, une reconnaissance des contraintes de terrain, et la mise en dialogue des acteurs.

Cette thèse propose ainsi une lecture située mais transposable de la fabrique de la science ouverte à l'échelle locale. Elle souligne l'importance d'un accompagnement contextualisé, relationnel et progressif, à même de faire évoluer les usages sans les contraindre. Elle rappelle enfin le rôle structurant des sciences de l'information et de la communication dans l'analyse et l'accompagnement des mutations info-documentaires contemporaines, à l'interface entre infrastructure technique, enjeux politiques et pratiques professionnelles. L'apparition récente du dispositif HALiance, proposant, entre autre, la suggestion automatique de dépôts à partir de sources tierces, marque une évolution significative dans la logique d'auto-archivage. En réduisant la part d'initiative individuelle, ce type d'outil peut bouleverser les représentations et les pratiques liées au dépôt scientifique. L'étude de son appropriation réelle par les chercheurs, ainsi que de ses effets sur la qualité, la diversité ou la temporalité des dépôts, constitue une piste féconde pour de futurs travaux. Elle permettrait de mieux cerner dans quelle mesure l'automatisation de l'accompagnement modifie non seulement les modalités d'accès, mais également les logiques d'adhésion à la science ouverte. En ce sens, l'étude des évolutions récentes des pratiques d'archivage peut, à nouveau, permettre de mieux comprendre les transformations en cours, à l'aune des dernières innovations.

Table des matières

Introduction.....	1
Positionnement et choix rédactionnels.....	2
Partie 1 - Histoire, développements et impacts du libre accès.....	5
1.1. Des échanges épistolaires à la crise des périodiques.....	6
1.2. Le libre accès à l'information scientifique.....	8
1.2.1. Initiatives pionnières et premières archives ouvertes.....	8
1.2.2. Développement d'infrastructures pour l'interopérabilité.....	9
1.2.3. Formalisation internationale du libre accès.....	12
1.2.4. Multiplication des déclarations et positionnements institutionnels.....	13
1.2.5. Expansion des politiques nationales et institutionnelles.....	16
1.2.6. Impact des crises sur l'accélération du libre accès.....	28
1.2.7. Vers une généralisation de la science ouverte.....	30
1.2.8. Les voix contraires.....	34
1.2.9. Conséquences d'une appropriation massive.....	34
Initiatives pionnières.....	35
Formalisation et mondialisation du libre accès.....	36
Politiques en faveur du libre accès et consolidation des infrastructures.....	37
Vers une recherche « ouverte ».....	37
1.3. Les archives ouvertes.....	44
1.3.1. L'archive ouverte HAL.....	44
1.3.2. L'archivage en sciences de l'information et de la communication.....	49
1.3.3. Les métadonnées.....	51
Partie 2 - SoVisu, un dispositif d'accompagnement à l'auto-archivage.....	53
2.1. Contexte.....	54
2.2. Un premier prototype fonctionnel.....	55
2.3. Extension du développement dans une logique d'accompagnement.....	59
2.3.1. Initialisation du dispositif.....	60
2.3.2. La collecte et le traitement de l'information.....	60
Une entrée via les chercheurs.....	61
Une mesure de qualité des dépôts.....	62
Indexation.....	63
2.3.3. La question des laboratoires.....	64
2.3.4. L'interface utilisateur.....	66
Profil.....	69
Synthèses.....	70
Ressources.....	77
Exploration.....	77
Plus d'informations.....	78
2.4. Limites.....	79
2.5. Conclusion.....	80

Partie 3 - Cartographie des pratiques d'archivage sur HAL.....	83
3.1. Méthodologie.....	84
3.1.1. L'analyse quantitative.....	84
3.1.2. Objectif de recherche.....	86
3.1.3. Périmètre de recherche.....	87
3.1.4. Caractérisation des dépôts et déposants.....	89
Les niveaux de contribution.....	90
Les logiques de dépôt.....	91
Les types de contributeurs.....	92
3.1.5. Infrastructure et aspects techniques.....	94
3.1.6. Description des corpus.....	96
Corpus de notices HAL en sciences, techniques et médecine.....	96
Corpus de notices HAL en sciences humaines et sociales.....	97
Corpus de notices HAL en sciences de l'information et de la communication.....	97
Corpus de notices HAL tamponnées Université de Toulon.....	98
3.1.7. Accès aux données.....	98
3.2. Résultats.....	99
3.2.1. Les sciences, techniques et médecine.....	99
3.2.2. Discussion intermédiaire.....	116
3.2.3. Les sciences humaines et sociales.....	117
3.2.4. Discussion intermédiaire.....	132
3.2.5. Des disparités notables dans les pratiques d'archivage entre les disciplines des sciences, techniques et médecine et des sciences humaines et sociales.....	134
3.2.6. Les sciences de l'information et de la communication.....	134
3.2.7. Discussion intermédiaire.....	141
3.2.8. L'Université de Toulon.....	142
3.3. Discussion.....	145
Conclusions et perspectives.....	149
Bibliographie.....	154

Bibliographie

Ambrosi, A., Peugeot, V., Pimienta, D., & Collectif. (2005). *Enjeux de mots : Regards multiculturels sur les sociétés de l'information, Edition multilingue français-espagnol-anglais-portugais*. C&F Editions.

API HAL | API Archive Ouverte HAL. (s. d.). Consulté 19 juillet 2024, à l'adresse

<https://api.archives-ouvertes.fr/docs/search>

Arènes, C., Maurel, L., & Rennes, S. (2022). *Guide d'application de la Loi pour une République numérique pour les données de la recherche* [Report, Comité pour la science ouverte].

<https://doi.org/10.52949/31>

Article 30—LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)—Légifrance. (2016).

https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000033202841

Ashta, E., Augouvernaire, M., Caillet, C., Laffont, M., Okret, C., & Pinet, N. (2020, juin 22).

Résultats de l'enquête Archives Ouvertes Couperin 2019. Journées CasuHAL 2020 initialement prévues à Montpellier du 22 au 24 juin 2020. https://archivesic.ccsd.cnrs.fr/sic_02864717

Banon, Shay. (2024, août 29). *Elasticsearch is Open Source, Again*. Elastic Blog.

<https://www.elastic.co/fr/blog/elasticsearch-is-open-source-again>

Baruch, P. (2007). Open Access Developments in France : The HAL Open Archives System.

Learned Publishing, 20, 267-282. <https://doi.org/10.1087/095315107X239636>

Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H., & Heller, L. (2018). *Open Science Training Handbook* (Version 1.0). Zenodo. <https://doi.org/10.5281/zenodo.1212496>

Biasi, P.-M. de. (2003). Sciences : Des archives à la genèse. Pour une contribution de la génétique des textes à l'histoire des sciences. *Genesis (Manuscripts-Recherche-Invention)*, 20(1), 19-52.

<https://doi.org/10.3406/item.2003.1240>

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ Data-Driven Documents. *IEEE Transactions on*

- Visualization and Computer Graphics*, 17(12), 2301-2309.
- <https://doi.org/10.1109/TVCG.2011.185>
- Bouchard, A., & Boudry, C. (2024). *Utilisation et usages des identifiants numériques chercheurs en France. Synthèse de l'enquête nationale 2023*. Réseau des URFIST. <https://hal.science/hal-04537803>
- Boudry, C., & Bouchard, A. (2017). L'exemple de ResearchGate. *médecine/sciences*, 33, 647-652.
- Boukacem-Zeghmouri, C. (2023, mai 8). La réforme de l'évaluation de la recherche, « le monde d'après ». *Journées Open Science 2023*.
- Broudoux, É. (2017). Autorité scientifique et épistémique à l'épreuve de la mesure des citations. *Études de communication. langages, information, médiations*, 48, Article 48.
- <https://doi.org/10.4000/edc.6841>
- Bryant, R., Fransen, J., family=Castro, useprefix=true, given=Pablo, prefix=de, Helmstutler, B., & Scherer, D. (2021). *Research Information Management in the United States : Part 1—Findings and Recommendations*. <https://doi.org/10.25333/8HGY-S428>
- Budapest Open Access Initiative. (2002). *Budapest Open Access Initiative*. <https://www.budapestopenaccessinitiative.org/>
- Cacheux, S., de Blignières, A., Marchelli, M., & Sizaret, T. (2023). *Projet SCBU Toulon : Organiser un service d'appui à la recherche*. Enssib.
- Cardon, D. (2011). *Zoomer ou dézoomer ? Les enjeux politiques des données ouvertes*. <http://owni.fr/2011/02/21/zoomer-ou-dezoomer-les-enjeux-politiques-des-donnees-ouvertes/>
- CCSD. (2022, septembre 23). *Haliance : Un projet structurant et innovant pour HAL*. <https://www.ccsd.cnrs.fr/2022/09/haliance-un-projet-structurant-et-innovant-pour-hal/>
- CCSD. (2025, février 10). *Présentation de HAL Monitor*. <https://doi.org/10.60527/a0y3-vf13>
- Centre pour la Communication Scientifique Directe. (2021). *L'odyssée de HAL, 2001–2021* (N. Fargier, Éd.). Centre pour la Communication Scientifique Directe (CCSD). <https://www.ccsd.cnrs.fr/>

- Chanier, T. (2004). *Archives ouvertes et publication scientifique. Comment mettre en place l'accès libre aux résultats de la recherche ?*
- Charnay, D., & Michau, C. (2007, novembre). L'archive ouverte HAL. *JRES* 2007.
<https://hal.science/hal-01101888>
- Chartron, G. (2010). *Quelle redistribution du pouvoir dans l'espace documentaire académique ?*
- Chartron, G., & Schöpfel, J. (2017). Open access et Open science en débat. *Revue française des sciences de l'information et de la communication*, 11. <https://doi.org/10.4000/rfsic.3331>
- Chaudiron, S., & Ihadjadene, M. (2010). De la recherche de l'information aux pratiques informationnelles. *Études de communication. langages, information, médiations*, 35, Article 35.
<https://doi.org/10.4000/edc.2257>
- CNRS. (2019). *Feuille de route du CNRS pour la science ouverte*. Centre national de la recherche scientifique. <https://www.cnrs.fr/fr/la-feuille-de-route-du-cnrs-pour-la-science-ouverte>
- COAR Task Force on Supporting Multilingualism and non-English Content in Repositories. (2023). *Recommendations for Multilingual and Non-English Content in Repositories*. Confederation of Open Access Repositories (COAR).
<https://www.coar-repositories.org/news-updates/coar-releases-recommendations-for-multilingual-and-non-english-content-in-repositories/>
- Comité pour la science ouverte. (2022). *Appel de Paris sur l'évaluation de la recherche*.
<https://www.ouvrirlascience.fr/appel-de-paris-sur-levaluation-de-la-recherche/>
- Courbet, D. (2013). *Pour un développement raisonné des méthodes expérimentales en SIC. Quelques intérêts épistémologiques du pluralisme méthodologique*.
- Couzinet, V. (2009). *Dispositifs info-communicationnels : Questions de médiations documentaires*. HERMES.
- Couzinet, V., & Liquète, V. (2021). La médiation documentaire, à la confluence entre information et communication. In E. Delamotte (dir) (Éd.), *Recherches francophones sur les éducations aux médias, à l'information et au numérique : Points de vue et dialogues* (p. 104-129). Presses de l'Enssib. <https://hal.science/hal-03226423>

- Cuxac, P., Lamirel, J.-C., & Kieffer, N. (2019, novembre 5). *SKEEFT : Indexing method taking into account the structure of the document*. 15th International Conference on Webometrics, Informetrics and Scientometrics and 20th COLLNET meeting.
<https://inria.hal.science/hal-03179724>
- Dacos, M. (2016). Préface à l'édition française. In P. Suber, *Qu'est-ce que l'accès ouvert ?* (p. 7-18). OpenEdition Press. <http://books.openedition.org/oep/1686>
- Desrosières, A. (2008). Gouverner par les nombres : L'argument statistique II. In *Gouverner par les nombres : L'argument statistique II*. Presses des Mines.
<https://books.openedition.org/pressesmines/341>
- Di Bitetti, M. S., & Ferreras, J. A. (2017). Publish (in English) or perish : The effect on citation rate of using languages other than English in scientific publications. *Ambio*, 46(1), 121-127. <https://doi.org/10.1007/s13280-016-0820-7>
- Elastic/elasticsearch*. (2024). [Java]. elastic. <https://github.com/elastic/elasticsearch> (Édition originale 2010)
- Elastic/kibana*. (2024). [TypeScript]. elastic. <https://github.com/elastic/kibana> (Édition originale 2013)
- Elasticsearch-dump/elasticsearch-dump*. (2024). [JavaScript]. Elasticsearch Dump.
<https://github.com/elasticsearch-dump/elasticsearch-dump> (Édition originale 2013)
- European Commission. (2018). *Turning FAIR into reality : Final report and action plan from the European Commission expert group on FAIR data*. Directorate-General for Research and Innovation. <https://doi.org/10.2777/1524>
- European University Association. (2021). *Read & Publish contracts in the context of a dynamic scholarly publishing system*. European University Association.
<https://www.eua.eu/resources/publications/970:read-publish-contracts-in-the-context-of-a-dynamic-scholarly-publishing-system.html>
- Fecher, B., & Friesike, S. (2014). Open Science : One Term, Five Schools of Thought. In S.

- Bartling & S. Friesike (Éds.), *Opening Science : The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (p. 17-47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2
- Foucault, M. (1966). *Les mots et les choses : Une archéologie des sciences humaines*. Gallimard.
- Francoeur, J., Paquin, É., Fiorini, S., Meurs, M.-J., & Larivière, V. (2025, avril 17). *Multilinguisme et découvrabilité des contenus scientifiques : Le pari de la traduction automatique*. Chaire de recherche du Québec sur la découvrabilité des contenus scientifiques en français. <https://chairedecouvrabilite.openum.ca/2025/04/17/multilinguisme-et-decouvrabilite-des-contenus-scientifiques-le-pari-de-la-traduction-automatique/>
- Gallezot, G., Rossi, C., Chartron, G., & Noyer, J.-M. (2003). *Conception d'une Archive ouverte en SIC : Le sens de la technique*.
- Gorz, A. (2003). *L'Immatériel : Connaissance, valeur et capital / André Gorz*. Galilée. Paris. <https://bm-grenoble.fr/Default/doc/SYRACUSE/619000/l-immateriel-connaissance-valeur-et-capital-andre-gorz>
- Guédon, J.-C. (2008). *Accès libre, archives ouvertes et États-nations : Les stratégies du possible*.
- Gueguen, G. (2019). Metadata quality at scale : Metadata quality control at the Digital Public Library of America. *Journal of Digital Media Management*, 7(2), 115-126.
- Harnad, S. (1994). *A Subversive Proposal*. <https://www.arl.org/wp-content/uploads/2006/01/subversive-proposal-electronic-publishing-jun05.pdf>
- Hcéres. (2024). *Export des données de production scientifique de HAL pour les évaluations du Hcéres : Guide de l'utilisateur*. Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres). <https://appui-evaluation.hceres.fr/hal>
- Hey, T., Tansley, S., Tolle, K., & Gray, J. (2009). *The Fourth Paradigm : Data-Intensive Scientific Discovery*. <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive>

-scientific-discovery/

Institut de l'information scientifique et technique (Inist) - CNRS/UAR76. (2021). *Science ouverte (thésaurus)*. <http://data.loterre.fr/ark:/67375/TSO>

Jeangirard, E. (2019). *Baromètre français de la Science Ouverte* (Note Flash du SIES 17). Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.
<https://www.enseignementsup-recherche.gouv.fr/fr/barometre-francais-de-la-science-ouverte-47401>

Jeanneret, Y. (2007). *Y-a-t-il, vraiment, des technologies de l'information ? : NOUVELLE EDITION REVUE ET CORRIGEE* (Enlarged édition). PU SEPTENTRION.

Jeanneret, Y., & Souchier, E. (2005). L'énonciation éditoriale dans les écrits d'écran. *Communication & Langages*, 145(1), 3-15. <https://doi.org/10.3406/colan.2005.3351>

Jouët, J. (2000). Retour critique sur la sociologie des usages. *Réseaux. Communication - Technologie - Société*, 18(100), 487-521. <https://doi.org/10.3406/reso.2000.2235>

Kembellec, G., Fournier-S'Niehotta, R., & Cubaud, P. (2017). L'histoire du Cédric : Penser un dispositif archivistique en histoire des sciences. *Cahiers d'histoire du Cnam*, vol.07-08(1), 133-153.

Kibana – com&doc. (s. d.). Consulté 4 mars 2025, à l'adresse
<https://com-et-doc.fr/les-declinaisons-de-hal/module-statistiques-kibana/>

Kidd, T. (2004). *Scottish declaration on open access*.
<https://api.semanticscholar.org/CorpusID:152918546>

Le Deuff, O. (2011). *La formation aux cultures numériques. Une nouvelle pédagogie pour une culture de l'information à l'heure du numérique Auteur* (Première édition). FYP éditions.

Mahé, A., & Prime-Claverie, C. (2017a). 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, Article 11. <https://doi.org/10.4000/rfsic.3315>

Mahé, A., & Prime-Claverie, C. (2017b). 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*, 20(2-3), 79-96.
<https://doi.org/10.3166/dn.2017.00010>
- Marty, E., & Gallezot, G. (2020). L'analyse et la visualisation de données en sciences de l'information et de la communication : Vers une convergence des questionnements et des pratiques de recherche ? *Les Cahiers de la SFSIC*, 10.
<http://cahiers.sfsic.org/sfsic/index.php?id=669>
- Mathieu, V., Aubry, P., & Marchal, J. (2003, novembre). Single Sign-On open-source avec CAS (Central Authentication Service). *JRES* 2003.
- Max-Planck-Gesellschaft. (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. <https://openaccess.mpg.de/Berlin-Declaration>
- Mayère, A. (2010). Organizational Communication and the Renewed Issue in Information and Communication Technologies. *Management Communication Quarterly*, 24(4), 612-619.
<https://doi.org/10.1177/0893318910381474>
- Merzeau, L. (2013). L'intelligence des traces. *Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*, 1(59), p.115.
- Meyriat, J. (1981). Document, documentation, documentologie. In *Schéma et Schématisation*.
- Meyriat, J., & Miège, B. (2002). Le projet des SIC : De l'émergent à l'irréversible (fin des années 1960-milieu des années 1980). In R. Boure (Éd.), *Les origines des sciences de l'information et de la communication : Regards croisés* (p. 45-70). Presses universitaires du Septentrion. <https://doi.org/10.4000/books.septentrion.113483>
- Michel, B., & Chekib, V. (2015). *Les pratiques de recherche documentaire, de publication et de diffusion scientifique des productions de la recherche à l'Université Paris-Sud : Questionnaire à destination des chercheurs, enseignants-chercheurs et doctorants* [Research Report]. Université de Paris 11.
<https://hal.archives-ouvertes.fr/hal-01292693>
- Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. (2021, juillet).

- Deuxième Plan national pour la science ouverte. Généraliser la science ouverte en France 2021-2024.*
<https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>
- names-dataset : The python library to handle names* (Version 3.1.0). (2022). [Logiciel].
- Noyer, J.-M. (1995). *Les sciences de l'information : Bibliométrie, scientométrie, infométrie*. Presses universitaires de Rennes.
- O'Donnell, J. J., & Okerson, A. (1995). *Scholarly journals at the crossroads : A subversive proposal for electronic publishing*. Office of Scientific & Academic Publishing, Association of Research Libraries. <https://catalog.hathitrust.org/Record/003013520>
- Paquien-séguy, F. (2012). L'utilisateur et le consommateur à l'ère numérique. In G. Vidal (Éd.), *La sociologie des usages : Continuités et transformations* (p. 10-29). Hermès Sciences Publications.
<https://hal.science/hal-01611076>
- Perriault, J. (2008). *La logique de l'usage : Essai sur les machines à communiquer*. Editions L'Harmattan.
- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA : A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- Pontika, N., & Knoth, P. (2015). *Open Science Taxonomy* [Other]. FOSTER.
<https://www.fosteropenscience.eu/foster-taxonomy/open-science>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics : A manifesto*.
<https://zenodo.org/records/12684249>
- Prime-Claverie, C., & Pouchot, S. (2010). Archives ouvertes et banques de données commerciales : Quelle visibilité pour la recherche en sciences de l'information et de la communication ? *Actes du congrès 2010 de l'Association canadienne des sciences de l'information*,
http://www.cais-acsi.ca/conf_proceedings_2010_fr.htm.
<https://hal.science/hal-00593931>
- Publishing options | Nature*. (s. d.). Consulté 10 août 2024, à l'adresse
<https://www.nature.com/nature/for-authors/publishing-options>

- R, M. T., Parmar, M., A, S. A., & Asjola, V. (2019). Research Information Management Systems : A Comparative Study. In R. Bhardwaj & P. Banks (Éds.), *Research Data Access and Management in Modern Libraries* (p. 54-80). IGI Global Scientific Publishing.
<https://doi.org/10.4018/978-1-5225-8437-7.ch003>
- Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science : The case of Colombian researchers in biological sciences. *PLoS ONE*, 15(9), e0238372.
<https://doi.org/10.1371/journal.pone.0238372>
- Reymond, D. (2023, novembre). SoVisu+ (SVP) : Optimisation de la qualité des données dans l'Enseignement Supérieur et la Recherche grâce à l'Intelligence Artificielle. *Journée d'étude annuelle sur l'IA dans les métiers de l'information et de la documentation*.
<https://hal.science/hal-04474282>
- Reymond, D., & Bretel, H. (2024). “ Dans l'ESR français, c'est l'Abes qui, à travers IdRef, incarne le lien entre référentiels, documentation et recherche ”. *Arabesques*, 112, 14-15.
<https://doi.org/10.35562/arabesques.3825>
- Reymond, D., & Galliano, C. (2019). *Cartographie de l'expertise des chercheurs de l'Université de Toulon* (p. 20) [Intern report]. Université de toulon. <https://hal.archives-ouvertes.fr/hal-02643329>
- Reymond, D., & Lapôtre, R. (2023, septembre 15). SoVisu+ : Starting point and foundations of a national CRIS. *Semantic Web in Libraries*.
- Reymond, D., & Tabariès, A. (2024). Decentralized research data management : Introducing SoVisu+. *Procedia Computer Science*, 249, 216-223.
<https://doi.org/10.1016/j.procs.2024.11.067>
- Reymond, D., Tabariès, A., & Bara, L. (2022a). *Accompagnement et visualisations de la science ouverte (SoVisu)*. <https://hal.science/hal-04103412>
- Reymond, D., Tabariès, A., & Bara, L. (2022b). *Open Science Visualisations (SoVisu)* [Logiciel].
<https://hal.archives-ouvertes.fr/hal-04104161>

- Schmitt, E. (2018). *Explorer, visualiser, décider : Un paradigme méthodologique pour la production de connaissances à partir des big data* [Phdthesis, Université de technologie de Compiègne].
<https://theses.hal.science/tel-01960545>
- Schöpfel, J. (2020). L'usage de la plateforme HAL par des unités de recherche. *I2D - Information, données documents*, 3(3), 167-198.
- Schöpfel, J., Chaudiron, S., Jacquemin, B., Kergosien, E., Prost, H., & Thiault, F. (2023). *The Transformation of the Green Road to Open Access* (2023020268). Preprints.
<https://doi.org/10.20944/preprints202302.0268.v1>
- 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. <https://doi.org/10.3917/i2d.222.0150>
- Schöpfel, J., Kergosien, É., Prost, H., Jacquemin, B., & Thiault, F. (2024). *L'utilisation de HAL. Pratiques, avis et alternatives*. <https://hal.science/hal-04523829>
- Simonnot, B. (2012). *L'accès à l'information en ligne : Moteurs, dispositifs et médiations*.
- Sociétés de la connaissance : Fractures et évolutions. (2009). In M. Durampart (Éd.), *Sociétés de la connaissance : Fractures et évolutions*. CNRS Éditions.
<https://books.openedition.org/editionscnrs/14364>
- Springer goes head to head with Elsevier on open access. (2004). *Interlending & Document Supply*, 32(4). <https://doi.org/10.1108/ilds.2004.12232dab.019>
- 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.
<https://doi.org/10.4000/rfsic.12258>
- Tabariès, A. (2024a). *Cartographie des pratiques d'archivage sur HAL* [Jeu de données].
<https://doi.org/10.34847/nkl.236023zx>
- Tabariès, A. (2024b). *Cartographie des pratiques d'archivage sur HAL - Archive ouverte HAL* [Logiciel].
<https://hal.science/hal-04654313>

- Tabariès, A., & Reymond, D. (2023, octobre). Analyse et situation des sciences de l'information et de la communication dans les pratiques d'archivage. *CiDE 23. 23ème édition du Colloque International sur le Document Electronique. Document et archivage : pratiques formelles et informelles dans les organisations*. <https://hal.science/hal-04230475>
- Varmus, H., Brown, P., & Eisen, M. (2001). *Public Library of Science Open Letter*. Public Library of Science. <https://plos.org/open-letter>
- Veissier, L. (2019, décembre 6). Franck Laloë : « On me soupçonnait de vouloir créer un big brother ». *TheMetaNews*.
<https://themeta.news/franck-laloe-on-me-soupconnaait-de-vouloir-creer-un-big-brother/>
- Wakim, N. (2025, mai 27). *Climat, biodiversité : Comment se passer de pesticides ?* Le Monde.
<https://podcasts.lemonde.fr/chaleur-humaine/202505270500-climat-biodiversite-comment-se-passer-de-pesticides>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zacklad, M. (2004). *Processus de documentarisation dans les Documents pour l'Action (DopA) : Statut des annotations et technologies de la coopération associées (nouvelle version corrigée)*.

Alaric Tabariès

IMSIC, Université de Toulon

Augmenter la découvrabilité scientifique : accompagner l'auto-archivage dans HAL ?

Résumé en français

Cette thèse interroge les pratiques d'auto-archivage scientifique dans le contexte français, à travers une analyse approfondie de l'archive ouverte HAL. Elle articule une approche socio-technique et informationnelle pour examiner les dynamiques de dépôt, la qualité des métadonnées et les logiques de découvrabilité associées. S'appuyant sur une recherche-action menée à l'Université de Toulon et sur le développement du dispositif SoVisu, ce travail met en lumière les tensions entre prescriptions institutionnelles et pratiques effectives, ainsi que les spécificités disciplinaires, notamment en sciences de l'information et de la communication. Il propose une typologie des contributeurs et des logiques de dépôt, tout en explorant les enjeux de visibilité, de traçabilité et d'évaluation dans le cadre de la science ouverte. En croisant analyse quantitative massive et réflexivité méthodologique, cette recherche contribue à une meilleure compréhension des conditions d'appropriation des archives ouvertes et de leur rôle dans l'écosystème de la communication scientifique contemporaine.

Mot clés : auto-archivage ; archives ouvertes ; découvrabilité ; science ouverte ; HAL ; métadonnées ; visualisation ; communication scientifique ; information scientifique et technique ; pratiques informationnelles.

Enhancing scientific discoverability: supporting self-archiving in HAL?

Résumé en anglais

This thesis investigates self-archiving practices within the French academic context by focusing on the open archive HAL. Adopting a socio-technical and informational approach, it explores deposit dynamics, metadata quality, and the underlying logics of discoverability. Grounded in an action research project conducted at the University of Toulon and centered on the development of the SoVisu platform, this work highlights the tensions between institutional guidelines and actual practices, with particular attention to disciplinary specificities in information and communication sciences. It proposes a typology of contributors and deposit logics, while examining issues of visibility, traceability, and evaluation within the broader framework of open science. By combining large-scale quantitative analysis and methodological reflexivity, this research provides a deeper understanding of how open archives are appropriated and their role in shaping contemporary scholarly communication.

Keywords : self-archiving ; open archives ; discoverability ; open science ; HAL ; metadata ; visualization ; scholarly communication ; scientific and technical information ; information practices.