

Sites de dépôt en libre accès et formes de médiations : quelles évolutions ?

Camille Prime-Claverie, Annaïg Mahé

► **To cite this version:**

Camille Prime-Claverie, Annaïg Mahé. Sites de dépôt en libre accès et formes de médiations : quelles évolutions ?. DocSoc 2013 - Diversification et renouvellement des médiations - Zagreb, Apr 2013, Zagreb, Croatie. pp.1-13, 2013. <sic_00866225>

HAL Id: sic_00866225

https://archivesic.ccsd.cnrs.fr/sic_00866225

Submitted on 26 Sep 2013

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.

Sites de dépôt en libre accès et formes de médiations : quelles évolutions ?

Open access repositories and forms of mediations : how do they evolve ?

Camille Prime-Claverie - MCF - Université Paris Ouest Nanterre La Défense - Dispositifs d'Information et de Communication à l'Ere Numérique – Paris, Ile-de-France (DICEN-IDF EA4420), camille.claverie@u-paris10.fr

Annaïg Mahé - MCF - Urfist de Paris / Ecole nationale des chartes - Dispositifs d'Information et de Communication à l'Ere Numérique – Paris Ile de France (DICEN-IDF EA4420), annaig.mahe@enc.sorbonne.fr

Résumé

Cette communication a pour ambition d'étudier les différentes formes de médiations mises en place autour des archives ouvertes en se centrant plus particulièrement sur le processus de dépôt des articles. Elle présente les résultats d'une étude quantitative réalisée à partir du corpus HAL pour la thématique des sciences de la vie sur une période de dix ans. Elle est fondée sur un modèle théorique exposant trois formes de médiation différenciées selon les acteurs concernés par le dépôt. L'objectif de cette recherche est de déterminer dans quelles proportions les logiques mises en œuvre dans ce corpus se rapprochent ou s'éloignent du modèle théorique proposé.

Mots-clés : libre accès, archives ouvertes, logiques de dépôt, médiation

This communication deals with the various forms of mediation about open archives with a focus on the process of article deposit. It shows the results obtained through a quantitative study realized from the HAL corpus for life sciences on a ten-year's period. It is based on a theoretical model presenting three forms of mediation differentiated according to the actors of the deposit. The aim of this research is to determine how close to or how separate from the theoretical proposed model the logics involved in this corpus are getting.

Keywords : open access, open archives, deposit logics , mediation

Introduction

Dans le contexte du libre accès à la communication scientifique, cette communication a pour objectif d'étudier les différentes formes de médiations mises en place autour des archives ouvertes. La notion de médiation recouvre de nombreuses acceptions, nous l'entendons ici comme le fait de «mettre en relation des ressources et des usagers », en nous focalisant sur les processus de dépôt des articles et les acteurs concernés par ces processus. Il s'agira d'analyser les différentes logiques de médiation qui se superposent dans les archives ouvertes et d'illustrer au moyen de données quantitatives quelles sont leur place et leur évolution depuis la création de ces dispositifs.

1. Contexte : Développement des sites de dépôt en libre accès

A partir du début des années 90, les tout premiers sites de dépôt d'articles scientifiques en libre accès étaient des sites thématiques, initiés par des chercheurs : the Mathematical Physics Preprint Archive (mp_arc) en physique mathématique, ArXiv (initialement xxx.lanl.org ou The Los Alamos Preprint Archive) en physique puis pour des disciplines connexes, RePEc (initialement WoPEc, Working Papers in Economy) en économie ou encore ArchiveSic pour les sciences de l'information et de la communication en France. Ces sites avaient pour objectif principal de développer une communication

scientifique directe (CSD) en permettant un accès libre, direct, aux résultats récents de la recherche par la mise à disposition des productions scientifiques (articles validés, preprints, ouvrages ou chapitres, rapports, etc.).

A partir des années 2000, et notamment suite à l'Initiative de Budapest pour le libre accès à la recherche en 2002 qui recommande le dépôt des articles dans les archives ouvertes (voie verte) et la publication dans des revues en libre accès (voie or)¹, les enjeux autour des archives ouvertes se sont institutionnalisés : un nombre croissant d'organismes de recherche se positionnent en faveur du libre accès, et surtout développent des sites de dépôt institutionnels dans le but de promouvoir la diffusion de la production scientifique de leurs communautés de recherche. Parallèlement, des politiques d'obligation de diffusion des articles scientifiques en libre accès (voie verte et/ou or) voient le jour, en premier lieu de la part d'organismes financeurs de la recherche, tels que le Wellcome Trust au Royaume-Uni ou les National Institutes of Health (NIH) aux Etats-Unis. Récemment, on a assisté à la mise en défaite d'une initiative de lobbying au Congrès américain visant à contrer ces politiques institutionnelles d'obligation de libre accès aux résultats de la recherche publique², et à un renforcement de la politique du Wellcome Trust³ comme de celle des NIH⁴, ces événements indiquant à eux seuls une montée en charge significative des politiques institutionnelles en faveur du libre accès, répercutée par d'autres prises de position officielles partout dans le monde (Commission européenne, Australie, UK Research Councils, etc.).

Des études récentes indiquent que le taux moyen de disponibilité des articles scientifiques en libre accès (voie verte et voie or), toutes disciplines confondues, est de 20 à 24% de la littérature publiée, avec de fortes variations disciplinaires [Gargouri et al., 2012 ; Björk et al., 2010]. Afin d'optimiser ce taux de remplissage des sites institutionnels, et parallèlement au développement des politiques institutionnelles d'obligation de dépôt, des conventions sont également passées entre des organismes producteurs ou financeurs de la recherche et des plateformes de publication et de diffusion mettant en place des procédures automatisées de transfert de documents de ces plateformes vers les sites institutionnels⁵.

Ces évolutions, d'initiatives disciplinaires à des développements institutionnels, traduisent, dans un même mouvement vers le libre accès à l'information scientifique, des intérêts différenciés selon les acteurs concernés, intérêts qui se superposent et ne se recouvrent pas toujours selon que l'on soit un chercheur (auteur/lecteur), un professionnel de l'information, une institution ou un organisme financeur de la recherche.

2. Formes de médiations autour des sites de dépôt

Si l'on reprend l'historique du développement du libre accès sous l'angle de la médiation, on constate que les modalités de mise à disposition des ressources sur les sites de dépôt révèlent aussi cette diversité et cette complexité. Nous nous intéressons ici précisément à la forme initiale de la médiation que constitue la mise à disposition du document sur le serveur. Qui en sont les acteurs, quels types de ressources mettent-ils à disposition et comment ? Ce sont les questions auxquelles nous avons cherché à répondre afin de mieux saisir les logiques à l'œuvre.

¹ Voir le texte de la déclaration de Budapest : <http://openaccess.inist.fr/?Initiative-de-Budapest-pour-l> (version en français) ainsi que le « Tour d'horizon de l'Open Access » sur le site d'ORBI, la DigiThèque de l'Université de Liège, <http://orbi.ulg.ac.be/project?id=01> qui présente, entre autres, les deux voies du libre accès.

² Jennifer Howard, Legislation to Bar Public-Access Requirement on Federal Research Is Dead, The Chronicle of Higher Education, 27 février 2012, <http://chronicle.com/article/Legislation-to-Bar/130949/>

³ Communiqué de presse, 28 juin 2012, <http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm>

⁴ Annonce de Sally Rockey, NIH Deputy Director, 16 novembre 2012, <http://nexus.od.nih.gov/all/2012/11/16/improving-public-access-to-research-results/>

⁵ Un exemple de ce type de conventions : Elsevier transfère, après une période d'embargo de 12 mois, les articles issus des financements des NIH vers PubMed Central.

Le contexte présenté précédemment nous permet de proposer un modèle, qui est une typologie reflétant les grandes phases historiques d'évolution du mouvement des archives ouvertes : l'élaboration d'une communication scientifique directe, l'implication institutionnelle et l'automatisation des processus. Ce modèle se décline en trois formes de médiation différenciées selon les acteurs concernés par le dépôt (auteur, tiers appartenant à l'institution, plateformes extérieures).

Le « niveau 0 » de la médiation : du chercheur au chercheur

Depuis l'origine, l'archive ouverte apparaît elle-même comme un dispositif de médiation, permettant une communication scientifique directe (CSD) entre les chercheurs fondée sur la mise à disposition du texte intégral du document scientifique récent. Ce modèle rappelle les premiers échanges épistolaires entre savants préexistants aux revues scientifiques. ArXiv, le serveur de preprints en physique et disciplines connexes en représente l'illustration la plus aboutie : l'objectif pour ce serveur étant de diffuser les résultats de la recherche avant même leur validation et leur publication dans une revue scientifique. A ce niveau, c'est l'auteur lui-même (ou un des auteurs) de l'article qui est le déposant : le dépôt se fait par et pour le chercheur.

Le « niveau 1 » de la médiation : des intermédiaires de l'institution au service de leurs communautés

En parallèle à la mise en place des sites de dépôt institutionnels, des services à valeur ajoutée se sont développés et des intermédiaires sont apparus dans la gestion du processus de dépôt : le déposant est alors un tiers, non auteur de l'article mais proche de sa communauté. Il appartient à l'institution d'au moins un des auteurs. Il peut s'agir d'un chercheur « référent » ou, plus souvent, d'un professionnel de l'information qui prend en charge le dépôt pour les auteurs et le valorise (ajout de métadonnées, gestion d'une collection par exemple) de façon manuelle ou automatisée. A ce niveau, la médiation autour du dépôt est envisagée comme un service pour le chercheur.

Le « niveau 2 » de la médiation : des intermédiaires extérieurs au service des institutions

Au fur et à mesure que la dimension politique du libre accès prend de l'ampleur, une nouvelle forme de médiation se développe autour de la gestion du dépôt. Les prises de position de grands organismes financeurs de la recherche ou la mise en place de protocoles institutionnels se concrétisent de plus en plus dans des procédures de transferts de métadonnées et/ou de documents entre éditeurs et sites de dépôt, voire d'un site de dépôt à un autre. A ce niveau de médiation, l'auteur n'est présent qu'au titre de métadonnées et non plus comme acteur ou bénéficiaire direct du dépôt.

Dans le cadre de cette analyse, nous cherchons à confronter ce modèle théorique de développement des archives ouvertes à un corpus spécifique afin d'analyser dans quelles proportions les logiques mises en œuvre dans ce corpus se rapprochent ou s'éloignent du modèle.

3. Corpus et méthodologie

3.1. Présentation du corpus

L'étude de terrain a été menée à partir de données extraites du site de dépôt HAL. Cette plateforme nous intéresse particulièrement du fait de ses spécificités organisationnelles et des motivations de dépôts contrastées qui en découlent. Alors que plus de 80% des archives ouvertes à travers le monde sont des archives institutionnelles⁶, HAL créée en 2001 par le Centre pour la communication scientifique directe (CCSD, CNRS) représente une spécificité française : c'est une archive nationale centralisée multidisciplinaire mais multi-interfaces, c'est-à-dire proposant différents portails permettant des accès à l'information par institutions, disciplines, projets, etc.

⁶OpenDOAR, 15 janvier 2013, <http://www.opendoar.org/index.html>

Historiquement, le modèle sur lequel est basé HAL est celui de la « communication scientifique directe », définie par Franck Laloé, un des premiers responsables du CCSD comme « une approche essentiellement « chercheur » : des archives alimentées directement par les chercheurs avec comme vocation première, la réalisation d'un outil scientifique offrant l'accès au document intégral ».

Rapidement, le site apparaît aussi comme un outil intéressant pour l'administration et la gestion de la recherche grâce à la collecte de métadonnées. A partir de 2006, la signature d'un protocole d'accord en fait le site de dépôt unique pour un certain nombre d'organismes de recherche français⁷. C'est également, au niveau disciplinaire, un site à la portée internationale grâce à une interconnexion avec d'autres sites de dépôts comme ArXiv pour la physique, PubMed Central pour les sciences biomédicales ou RePec pour l'économie.

Au niveau international, les études menées sur les usages des archives ouvertes par les chercheurs sont essentiellement multidisciplinaires et basées sur des données d'enquêtes issues de questionnaires cherchant à en analyser les perceptions et les pratiques [Nicholas et al., 2012 ; Fry et al, 2011 ; Swan et Brown, 2005]. Pour cette étude, nous avons adopté une démarche quantitative et choisi de travailler à partir des dépôts effectués pour une discipline. Prenant en compte les spécificités de Hal énoncées précédemment, l'objectif de cette étude est d'examiner comment se sont mises en place et comment ont évolué différentes formes de médiations autour des modalités de dépôt depuis la mise en place de l'archive, c'est-à-dire sur une période de dix ans.

En mai 2012, le moissonnage de l'archive ouverte HAL par le protocole OAI-PMH pour le sous-ensemble thématique des sciences de la vie nous a permis de rassembler un ensemble de 53.081 notices au format XML selon le standard de métadonnées proposé par HAL (OAI-HAL)⁸. Ce jeu de données a ensuite été complété en mars 2013 par l'ensemble des dépôts effectués en 2012 pour cette même thématique. Le corpus final comporte 58.692 notices.

Le standard de métadonnées proposé par HAL, plus détaillé que le Dublin Core, offre l'avantage de mentionner de manière précise chacun des auteurs et leurs affiliations et d'indiquer qui est le déposant des notices. Ceci nous permettra par la suite d'identifier qui opère le dépôt d'un document.

3.2. Méthodologie

Nous avons développé une méthodologie semi-automatisée permettant de déterminer qui a effectué un dépôt : l'un des auteurs, un contributeur tiers de l'institution (documentaliste, chercheur référent, etc.), ou un intermédiaire extérieur (autre plateforme par exemple). Cette méthode s'appuie sur un script qui procède par comparaison de chaînes de caractères entre champs auteur et contributeur. Pour tenter de repérer dans notre corpus chacun des trois niveaux de médiations préalablement décrits, nous avons défini un certain nombre de critères les caractérisant.

Critères généraux :

- qui est le contributeur (déposant) : auteur, intermédiaire de l'institution, intermédiaire extérieur ;
- âge du document référencé ;
- présence du texte intégral ou notice bibliographique seule.

Critères pour chaque niveau :

Niveau 0 : HAL (et ses portails) comme dispositif de médiation assurant une communication scientifique directe et immédiate (du chercheur au chercheur)

⁷Protocole d'accord en vue d'une approche coordonnée, au niveau national, pour l'archivage ouvert de la production scientifique, http://www.couperin.org/IMG/pdf/protocoleAO_pour_signature.pdf

⁸Requête formulée pour le moissonnage des notices : http://hal.archives-ouvertes.fr/oai/oai.php?verb=ListRecords&metadataPrefix=oai_hal&set=SDV

- présence du texte intégral ;
- dépôt effectué par un auteur du document ;
- dépôt récent (preprint ou dépôt effectué l'année de publication du document ou l'année suivante).

Niveau 1 : des intermédiaires de l'institution au service de leurs communautés

- Dépôt (notice bibliographique accompagnée ou non du texte intégral) déposée par un tiers (non auteur du document) appartenant à l'institution d'un des auteurs.

Niveau 2 - des intermédiaires extérieurs au service des institutions

- transfert massif de notices avec ou sans texte intégral à partir d'une autre plateforme.

4. Résultats

4.1. Types d'enregistrements en sciences de la vie sur Hal

Les références collectées ont été déposées entre 2002 et 2012. Elles représentent presque 10% des références déposées dans HAL sur la même période.

Parmi les quinze types de documents déposés pour l'ensemble de notre corpus, les quatre principaux sont : article publié dans une revue à comité lecture (79%), communication avec actes (5%), thèse (5%), communication sans actes (4%). On constate d'emblée la place prépondérante des articles scientifiques qui constituent près des 4/5 de l'ensemble du corpus (Fig. 1).

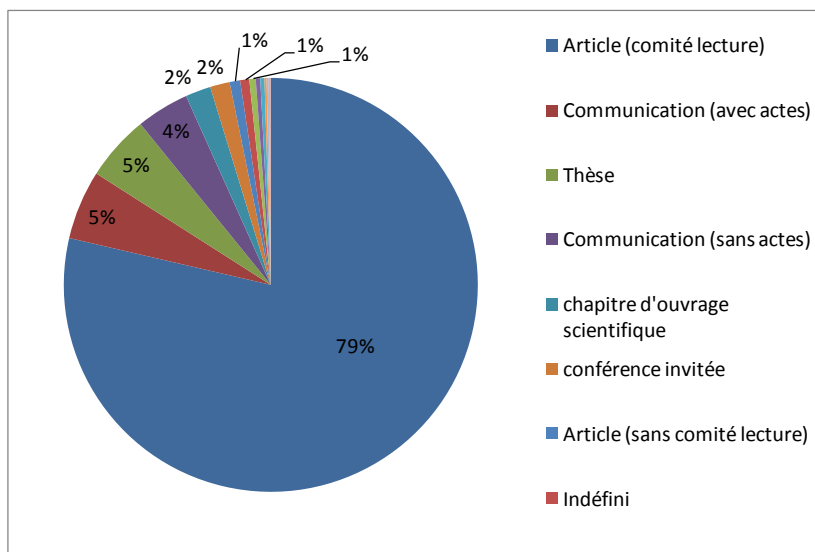


Figure 1 : Répartition des types de documents du corpus.

Les trois premières années, le nombre de dépôts relativement faible concerne essentiellement des thèses. C'est surtout à partir de l'année 2006 que le nombre de dépôts s'intensifie et que le type de document se diversifie. Cette augmentation est concomitante de la signature du protocole d'accord inter-institutionnel et de l'ouverture de plusieurs sites de dépôts par des organismes de recherche, notamment en sciences de la vie : HAL-Inserm en janvier 2006 (26% du corpus), HAL-Pasteur (3%) en mars 2007, HAL-Inria (9%) en avril 2005 pour l'informatique (bio-informatique en sciences de la vie).

La grande majorité des enregistrements concerne des dépôts de notices bibliographiques seules (75,9% du corpus), ce qui signifie que moins d'un quart des enregistrements apparaissent donc avec un texte intégral associé. Ce taux augmente légèrement si on inclut les notices qui pointent vers une plateforme intégrant l'accès libre au document en texte intégral (PubMed Central et ArXiv notamment).

4.2. Des modalités de dépôts différenciées selon le type de contributeur

Pour l'ensemble du corpus, les dépôts émanant de contributeurs tiers appartenant à l'institution de l'un des auteurs sont majoritaires : ils sont responsables de près de 71% des enregistrements, tandis que les dépôts effectués par l'un des auteurs du document en représentent 26% et ceux effectués par des intermédiaires extérieurs presque 3% (Fig. 2).

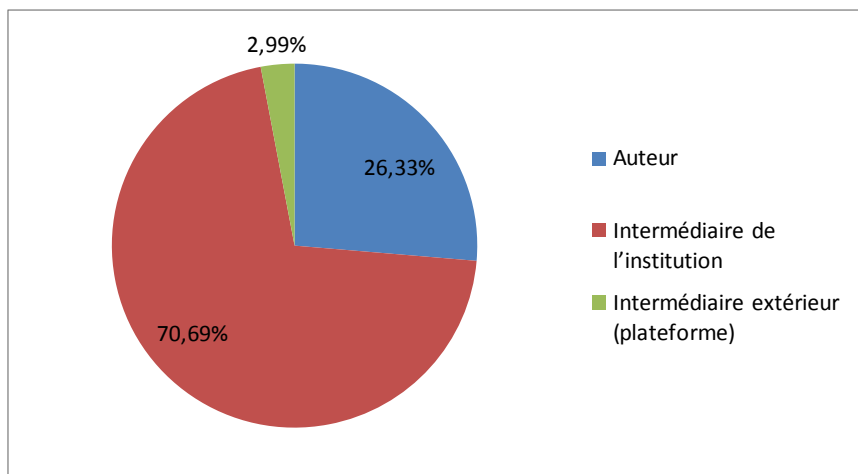


Figure 2 : Répartition des dépôts en fonction du type déposant.

L'évolution des dépôts selon les contributeurs (Fig.3) illustre la place prépondérante que prennent les intermédiaires de l'institution à partir de 2006, et l'arrivée de nouveaux acteurs, les intermédiaires extérieurs, très marginalement en 2010, et s'intensifiant par la suite.

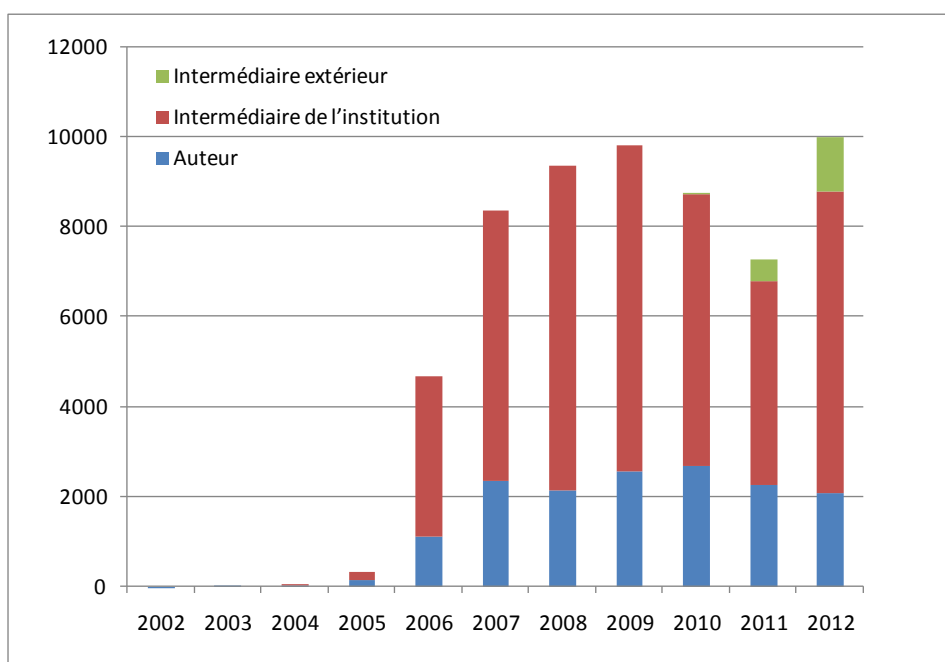


Figure 3 : Evolution des dépôts en fonction du type déposant.

Une analyse plus fine des types de dépôts selon le type de contributeur met bien en évidence des pratiques de dépôt différenciées selon que le contributeur est l'un des auteurs du document, un intermédiaire de l'institution ou un intermédiaire extérieur. La figure 4 met en exergue la prépondérance des dépôts de notices bibliographiques seules (sans texte associé) effectués par les intermédiaires de l'institution (presque 86 % des dépôts effectués par ceux-ci). Les auteurs déposent de manière presque équilibrée des notices seules ou des notices accompagnées du document en texte

intégral. Les intermédiaires extérieurs, quant à eux déposent exclusivement des références accompagnées du document complet.

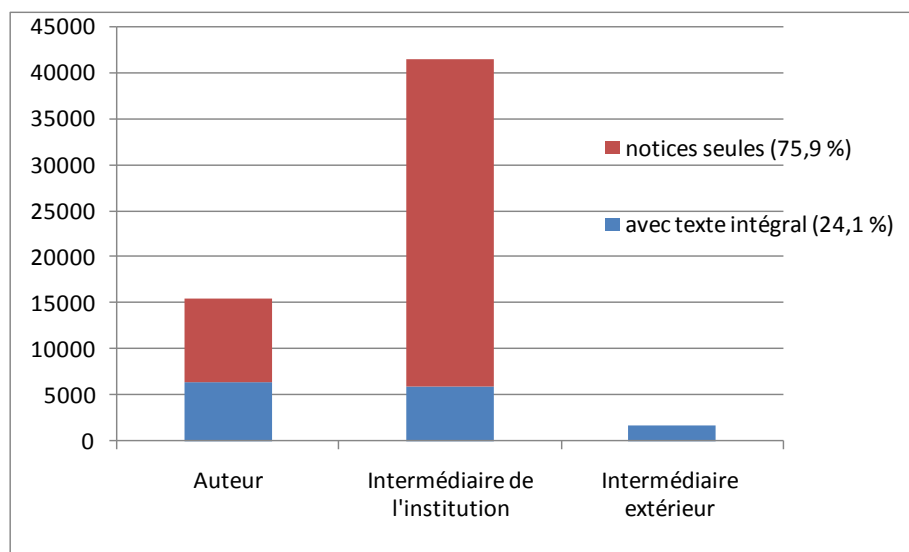


Figure 4 : Type de dépôts en fonction du type de déposant.

4.3. Du modèle au terrain : les trois niveaux de médiation

Si on analyse plus en détail les types d'enregistrements effectués selon le type de contributeur, on constate un décalage entre le modèle proposé et la réalité du corpus en sciences de la vie et notamment sur l'utilisation de Hal pour la CSD. Le tableau I résume la variété des logiques de dépôts constatées pour chaque niveau de médiation, décrivant une réalité plus complexe que celle de notre typologie de départ.

		Auteur	Intermédiaire de l'institution	Intermédiaire extérieur
Notice avec texte intégral	Document récent	Communication scientifique directe (CSD)	Communication scientifique médiée (CSM)	
	Document ancien	Archivage		
Notice seule	Document récent	Visibilité du front de recherche (référencement)		
	Document ancien	Visibilité des recherches antérieures (recensement)		

Tableau I : Typologie des dépôts.

Nous décrivons ci-dessous le détail des résultats selon les types de contributeurs (Tableau II) afin d'analyser quelles sont les pratiques de dépôt en sciences de la vie et leur évolution dans le temps.

		Auteur	Intermédiaire de l'institution	Intermédiaire extérieur
Notice avec texte intégral	Document récent	7,08%	6,48%	2,20%
	Document ancien	3,99%	3,60%	0,79%
Notice seule	Document récent	7,47%	33,74%	0%
	Document ancien	7,80%	26,86%	0%

Tableau II : Répartition des logiques de dépôts.

Le dépôt par un auteur

Les résultats du Tableau II montrent que les dépôts par les auteurs se répartissent de manière assez équilibrée entre une logique de communication scientifique directe, telle que nous l'avons initialement définie, une logique de référencement (mise en visibilité du front de la recherche), et une logique de recensement des notices antérieures, correspondant respectivement à 7%, 7,5% et 7,8% des dépôts constatés dans l'ensemble du corpus. Une quatrième logique, plus marginale, se dégage, correspondant à une logique d'archivage de documents plus anciens (4%). Si on analyse l'évolution dans le temps de chacune de ces logiques (Fig. 5), on constate, sur les dernières années, une baisse assez forte du recensement (notices plus anciennes) et de l'archivage, tandis que la mise en visibilité de documents récents, avec ou sans texte intégral associé, augmente.

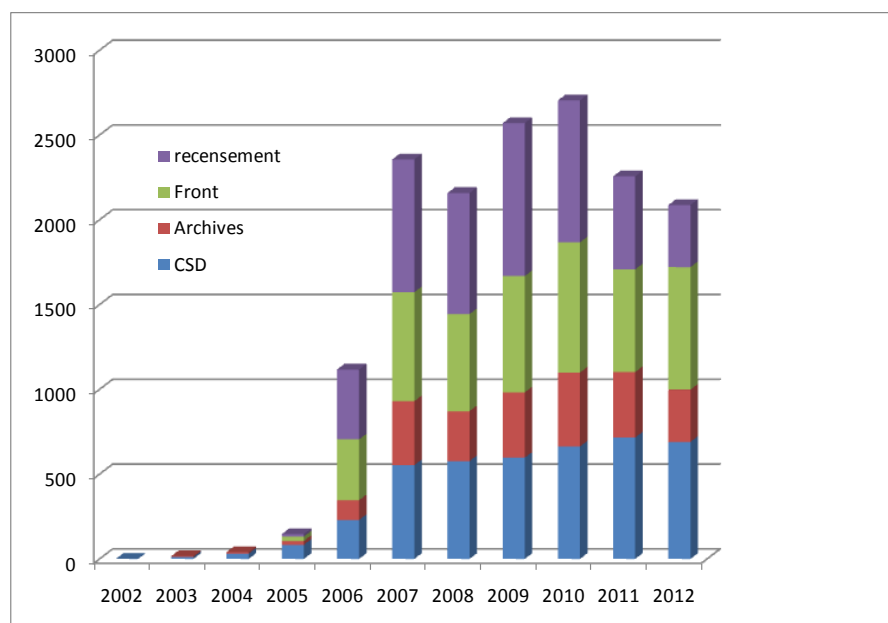


Figure 5 : Evolution des types de dépôts par les auteurs.

Le dépôt par un tiers de l'institution

Comme indiqué plus haut, les dépôts par un tiers de l'institution sont largement majoritaires, mais ce qui est particulièrement frappant, c'est la part des enregistrements sans aucun texte intégral qui

représente 60% du corpus. Parmi ces dépôts, c'est la logique de référencement du front de la recherche qui est légèrement majoritaire (33,7% du corpus). En ce qui concerne les dépôts avec un texte intégral associé (10% du corpus), les deux tiers (6,5% du corpus) correspondent à ce que nous avons défini comme une « communication scientifique médiée », c'est-à-dire donnant accès à des documents récents. Le dernier tiers des dépôts avec texte intégral correspond plutôt à une logique d'archivage de documents plus anciens. L'évolution dans le temps montre une grande irrégularité, quelque soit le type de dépôts (Fig.6).

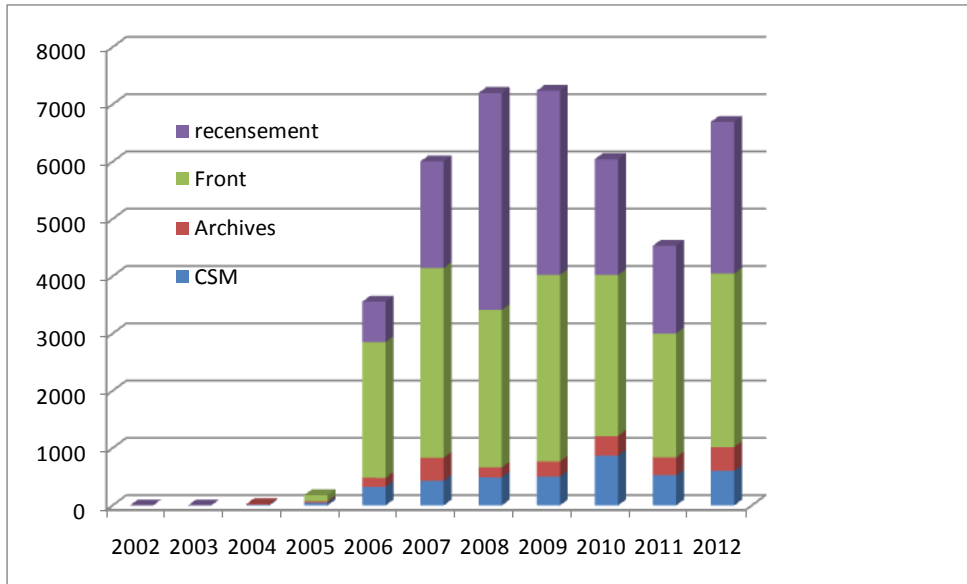


Figure 6 : Evolution des types de dépôts par les intermédiaires de l'institution.

Le dépôt par un intermédiaire extérieur

Les dépôts de documents en texte intégral effectués par des intermédiaires extérieurs à partir de 2010 concernent en grande majorité des documents récents (Fig. 7). Ces dépôts, en forte croissance depuis deux ans, montrent la place prépondérante prise par ces nouveaux acteurs dans le paysage des archives ouvertes. En 2012, ils deviennent les fournisseurs les plus importants de documents en texte intégral (Fig. 8).

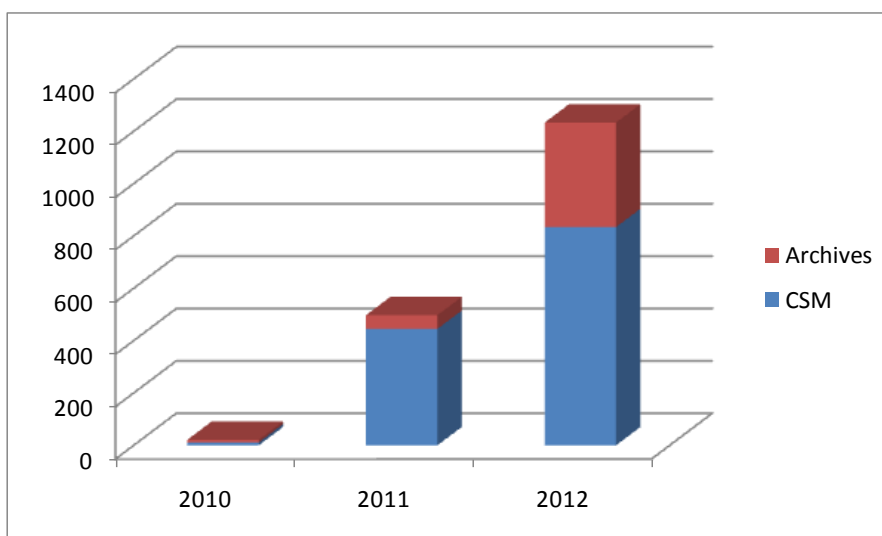


Figure 7 : Evolution des types de dépôts par les intermédiaires extérieurs.

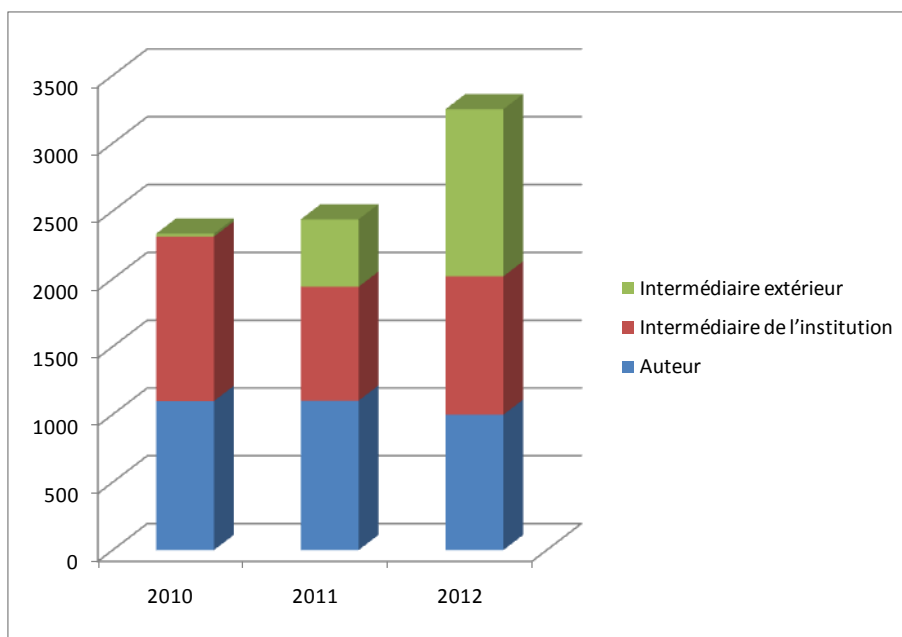


Figure 8 : Répartition des dépôts en texte intégral.

5. Discussion

Les résultats de cette première analyse de données collectées pour les dépôts sur HAL en sciences de la vie apportent un certain nombre d'éclairages sur les logiques d'utilisation d'un tel site de dépôt de documents scientifiques ainsi que sur les formes de médiation qui s'y développent.

Tout d'abord, on constate que les dépôts effectués sur HAL reflètent bien la spécificité du mode de communication scientifique dominant en sciences de la vie, les articles scientifiques représentant la plus grande part des enregistrements effectués sur HAL (notices bibliographiques seules ou avec le texte intégral associé).

Cependant, on note également que la logique de la communication scientifique directe telle que nous l'avons définie en amont, n'est que très partiellement atteinte en ce qui concerne ce domaine. En effet, si l'on considère une définition « stricte » de la CSD (dépôt d'un document récent par l'auteur), moins d'un enregistrement sur 15 y correspond (7% du corpus). Cependant, si on élargit cette définition en y incluant tous les dépôts avec texte intégral par un auteur (récent ou plus ancien), on arrive à près d'un document sur 9 (11.1% du corpus), et un peu moins d'un document sur 4 (24,1 % du corpus) en considérant la totalité des dépôts avec texte intégral. L'accès libre au texte intégral des documents récents n'est donc effectif que pour moins d'un quart du corpus total des enregistrements, et moins de la moitié de ces dépôts seulement est effectuée par les auteurs eux-mêmes.

Malgré ce taux de CSD globalement faible, on note la volonté des chercheurs de rendre visibles les travaux récents, avec ou sans texte intégral associé. Une explication du faible taux de dépôt des documents en texte intégral par les chercheurs provient peut-être de la nécessité dans ce domaine particulier de ne diffuser les textes qu'une fois validés et publiés. De ce fait, les périodes d'embargo définies par les éditeurs scientifiques les empêchent sans doute de déposer systématiquement le texte intégral avec la notice. Une étude de 2005 (Swan et Brown) sur les pratiques d'auto-archivage indique que les auteurs en sciences de la vie ont effectivement tendance à déposer largement plus de postprints (articles validés) que de preprints. De la même manière, une synthèse plus récente du groupe de travail sur le libre accès du Groupement français de l'industrie de l'information (GFII) fait la distinction entre deux modèles d'archives ouvertes : le modèle d'Arxiv, historiquement le premier qui met à disposition les articles avant soumission, et « le modèle de la préservation et de la valorisation de données préalablement validées par les pairs, très présent en sciences de la vie où la diffusion ne s'effectue qu'après la validation et la publication de l'article » (Synthèse GFII, 2010, p 38). Cette pratique

correspond également à la politique de l'archive des National Institutes of Health aux USA (NIH), PubMed Central, sur laquelle les articles mis librement à disposition sont au minimum préalablement validés.

Par ailleurs, l'étude de Brown et Swan précédemment citée montre que pour une majorité de chercheurs l'âge moyen des articles utilisés peut remonter jusqu'à dix ans (avec quelques variations disciplinaires, mais le constat est aussi valable pour les sciences de la vie). Cela permet donc d'élargir le principe de la communication scientifique directe en y ajoutant une logique d'archivage que l'on voit bien à l'œuvre ici.

En ce qui concerne l'irrégularité des dépôts effectués par les intermédiaires de l'institution, plusieurs raisons pourraient l'expliquer : il pourrait s'agir d'une mise à jour du recensement de la production scientifique dans le cadre du fait de campagnes d'évaluation (tous les quatre ans environ) ou d'un « effet de rattrapage » par le dépôt massif d'enregistrements de nouveaux entrants (institutions qui ouvrent un site de dépôt). Des analyses plus approfondies sur les motivations des déposants permettraient de mieux comprendre les logiques à l'œuvre.

Cette étude met également en exergue l'arrivée de nouveaux acteurs dans le paysage des archives ouvertes : les intermédiaires extérieurs. Dans ce corpus, il s'agit de l'« ABES STAR » et de BioMed Central, le premier étant l'Agence bibliographique de l'enseignement supérieur, l'acteur institutionnel de dépôt pour les institutions ayant choisi le dépôt de la version électronique pour les thèses soutenues dans l'établissement via le processus automatisé STAR, et le second étant un éditeur de revues en libre accès en sciences biomédicales qui transfère automatiquement sur HAL les métadonnées ainsi que le texte intégral des articles publiés dans ses revues par des chercheurs des institutions ayant choisi HAL comme portail de dépôt⁹. Ces dépôts automatisés sont non seulement récents (depuis 2010 pour l'Abes, et 2011 pour BioMed Central) mais ils représentent également une part importante, et croissante, des dépôts massifs puisqu'ils représentent près de 20% des dépôts de texte intégral en 2011 et près de 38% en 2012 (Fig. 8). Ces dépôts participent largement à l'augmentation du nombre de dépôts de texte intégral, et on assiste, dans ce cas, à la voie or qui se met au service de la voie verte.

Au vu de ces résultats, il est clair que l'évolution historique des modalités de dépôt sur les archives ouvertes, telle que synthétisée dans notre modèle à trois niveaux, ne correspond pas à la déclinaison disciplinaire d'implémentation des archives ouvertes. On savait déjà que la publication par la voie or est prépondérante dans les sciences de la vie (Björk, 2010), mais on constate aussi ici que le principe de CSD s'applique différemment de celui pratiqué dans les disciplines à l'origine des premiers sites de dépôt, notamment avec la logique d'archivage constatée.

Conclusion

Ces nuances dans les logiques de dépôt observées ici sur un corpus en sciences de la vie font écho à la complexité et la diversification du paysage naissant du libre accès à l'information scientifique que nous mentionnions au début de cet article, et cela montre aussi qu'il est non seulement impossible d'expliquer ce paysage de façon simpliste, mais qu'il est aussi nécessaire de mieux étudier les logiques spécifiques à des communautés particulières. Nous indiquions également en introduction que la plupart des études sur ce sujet sont multidisciplinaires et basées sur des questionnaires. Cette étude quantitative n'est qu'un début d'analyse des pratiques effectives et des logiques d'acteurs autour d'un site de dépôt de documents scientifiques, et il serait particulièrement intéressant de la prolonger soit en comparant ces résultats avec d'autres corpus disciplinaires, voire d'autres sites de dépôts, soit en la complétant par des analyses qualitatives permettant de mieux comprendre les logiques observées et leurs évolutions. Cela permettrait notamment aux auteurs, mais également aux acteurs censés les guider et les seconder, de mieux comprendre les opportunités ainsi que les limites liées au développement des questions du libre accès dans leurs communautés spécifiques.

⁹ C'est le cas notamment pour l'Inserm depuis avril 2011 : <http://www.hal.inserm.fr/>, mais aussi de l'Inria depuis novembre 2012 : <https://www.inria.fr/ist/adhesion-inria-a-bmc-biomed-central/>

Une autre question qui survient au vu de ces résultats est dans quelle mesure la voie or du libre accès viendrait se mettre au service de la voie verte, du moins pour les sciences de la vie, à moins qu'il ne s'agisse plus simplement d'une supplantation de la voie verte par la voie or, les chercheurs (via leurs institutions) faisant le choix premier de la publication dans des revues en libre accès, et laissant la mise en visibilité de leur production aux mains de tiers ? L'observation des évolutions à venir permettra certainement de distinguer les pratiques les plus répandues et adéquates selon les différentes communautés scientifiques.

Références

BJORK (B.-C.), WELLING (P.), LAAKSO (M.), MAJLENDER (P.), HEDLUND (T.), *Open Access to the Scientific Journal Literature: Situation 2009*, dans *PLoS ONE* 5(6), 2010. Consulté le 25 mars 2013.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0011273>

CIBER Research Ltd., *PEER Usage study. Final report*, juin 2012. Consulté le 15 janvier 2013.

http://www.peerproject.eu/fileadmin/media/reports/20120618_D5_2_PEER_Usage_Study_6_Month.pdf

FRY (J.), PROBETS (S.), CREASER (C.), GREENWOOD (H.), SPEZI (V.), WHITE (S.), *PEER behavioural research : authors and users vis-à-vis journals and repositories. Final report*, août 2011. Consulté le 15 janvier 2013.

http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf

GARGOURI (Y.), LARIVIERE (V.), GINGRAS (Y.), CARR (L.) et HARNAD (S.), *Green and Gold Open Access percentages and growth, by discipline. 17th International Conference on Science and Technology Indicators (STI), Montréal, CA, 05 - 08 Sep 2012*. Consulté le 25 mars 2013.

<http://eprints.soton.ac.uk/340294/>

GFII, *Groupe de travail sur le libre accès, document de synthèse*, janvier 2010. Consulté le 25 mars 2013.

<http://www.gfii.fr/uploads/docs/groupe-de-travail-gfii-sur-le-libre-acces-mise-en-ligne-de-la-synthese-des-discussions-et-des-recommandations.pdf>

NICHOLAS (D.), ROWLANDS (I.), WATKINSON (A.), BROWN D.), JAMALI (H. R.), *Digital repositories ten years on : what do scientific researchers think of them and how do they use them ?*, dans *Learned Publishing*, 25, 2012, pp. 195-206.

SWAN (A.), BROWN (S.), *Open access self archiving : an author study*, Key Perspectives, 2005. Consulté le 15 janvier 2013.

<http://cogprints.org/4385/1/jisc2.pdf>