



Ouvrir l'accès aux dernières publications de son établissement avec HAL

Maxence Larrieu, Dorothée Pain

► To cite this version:

Maxence Larrieu, Dorothée Pain. Ouvrir l'accès aux dernières publications de son établissement avec HAL : Retour d'expérience sur la mise en place d'un chantier d'import effectué à l'université de Versailles Saint-Quentin-en-Yvelines. Bulletin des Bibliothèques de France, 2021. hal-03339754

HAL Id: hal-03339754

<https://hal.science/hal-03339754>

Submitted on 10 Sep 2021

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

OUVRIR L'ACCÈS AUX DERNIÈRES PUBLICATIONS DE SON ÉTABLISSEMENT AVEC HAL

Retour d'expérience sur la mise en place d'un chantier d'import effectué à l'université de Versailles Saint-Quentin-en-Yvelines

Maxence Larrieu

Il a travaillé sur la science ouverte à l'université Paris-Est Marne-la-Vallée, à l'université d'Angers, où il a contribué au projet Conditor, et dernièrement à l'université de Versailles Saint-Quentin-en-Yvelines où il est chargé de la valorisation des publications. Il intervient à l'Enssib dans le cadre de formations sur la bibliométrie et l'accès ouvert.

Dorothée Pain

Elle est ingénieure d'études à l'université de Versailles Saint-Quentin-en-Yvelines (UVSQ), au sein de la Direction des bibliothèques et de l'information scientifique et technique (DBIST). Elle est responsable de la bibliothèque numérique de l'université et administre le portail HAL. Elle coordonne la cellule ebook du consortium Couperin et fait partie du bureau professionnel.

L'archive ouverte HAL a été conçue pour permettre aux chercheurs de partager leurs publications en accès ouvert. En 2016, la loi pour une République numérique (LRN) facilite les aspects juridiques de ce partage. Les conditions techniques et juridiques sont ainsi réunies et on aurait pu s'attendre à une augmentation nette du nombre de dépôts dans HAL. Néanmoins, rien de tel n'apparaît dans le *Baromètre français de la Science ouverte*¹, dans lequel ressort une augmentation progressive du pourcentage de publication en accès ouvert *via* archive, d'une moyenne de 0,8 % chaque année.

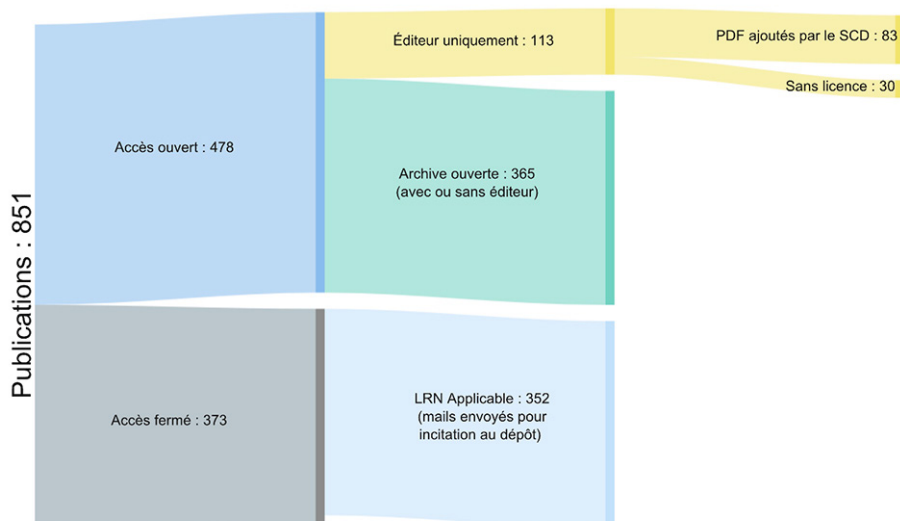
Le nombre d'articles déposé dans HAL directement par les auteurs représente moins de 20 % de l'ensemble des articles présents dans l'archive². Face à ce constat, le SCD de l'UVSQ reprend courant 2020 un chantier visant à augmenter l'accès ouvert aux publications en utilisant son portail HAL. Le présent article traite d'une solution

1 <https://ministeresuprecherche.github.io/bsol/> (consulté le 14 mars 2021)

2 Arrondi obtenu en utilisant directement l'Application Programming Interface (API) de HAL : voir la note 6 pour la méthode. La faible utilisation des archives ouvertes non disciplinaires n'est pas une spécificité française, mais un constat partagé par la communauté. Voir Andrés FERNÁNDEZ-RAMOS, Leticia BARRIONUEVO, « Value-added services in institutional repositories in Spanish public universities », *Information Research: an international electronic journal*, mars 2021, 26, n° 1. En ligne : <https://doi.org/10.47989/irpaper895>.

technique et de l'organisation d'un travail en équipe. Au total, 6 agents sont mobilisés, permettant de traiter une centaine de publications par mois. Depuis le commencement du chantier en mai 2020, 851 publications ont été traitées. Certaines étaient déjà présentes dans HAL (37 %) et plus de la moitié étaient déjà en accès ouvert (56 %).

■ **Figure 1. Statistiques des publications traitées**



L'AUTO-ARCHIVAGE DES PUBLICATIONS : ENTRE PRINCIPES ET RÉALITÉ

Avec plus de 3 millions de publications tous azimuts et environ 150 portails institutionnels³, HAL, à l'aune de ses 20 ans, apparaît fortement intégrée dans le paysage académique français. Toutefois, certains aspects de HAL soulèvent des questions. Relevons par exemple l'ambivalence entre les notices avec et sans texte intégral, ou encore le principe d'auto-archivage à l'origine de sa conception. Le fonctionnement de HAL — et des archives ouvertes en général — se base en effet sur l'auto-archivage, non pas dans le sens d'« automatique », mais dans celui où les auteurs eux-mêmes effectuent les dépôts des publications⁴. Néanmoins, plusieurs décennies après l'arrivée des archives ouvertes, force est de constater que l'auto-archivage reste limité.

3 L'enquête de 2019 sur les archives ouvertes réalisées par le consortium Couperin relève que neuf universités sur dix ont fait le choix d'un portail HAL. En ligne : <https://www.couperin.org/site-content/261-a-la-une/1407-resultats-de-l-enquete-sur-les-pratiques-de-publication-et-d-acces-ouvert-des-chercheurs-francais> (consulté le 14 mars 2021).

4 Le principe d'auto-archivage est à la genèse du mouvement de l'accès ouvert. Il est la première stratégie présentée, avant les revues en libre accès, dans l'Initiative de Budapest pour l'accès ouvert de 2002 : « Auto-archivage : en premier lieu, les savants ont besoin d'outils et d'assistance pour déposer leurs articles de revues à comité de lecture dans des archives électroniques ouvertes, une pratique communément appelée auto-archivage ». Traduction de la Budapest Open Access Initiative. En ligne : <https://www.ouvrirlascience.fr/initiative-de-budapest-pour-lacces-ouvert/> (consulté le 12 mars 2021). Voir également la préface de Marin DACOS dans Peter SUBER, *Qu'est-ce que l'accès ouvert ?* nouvelle édition, Marseille, OpenEdition Press, 2016, p. 11. En ligne : <http://books.openedition.org/oep/1600> (consulté le 12 mars 2021).

Une simple utilisation de l'API⁵ dans HAL montre que seulement 13 % des articles de revues⁶ sont déposés en texte intégral par les auteurs eux-mêmes. Ce constat est à mettre en regard de la LRN de 2016. L'article 30 facilite en effet l'auto-archivage en accordant le droit aux auteurs de déposer la version acceptée pour publication de leur article de revue⁷, même si ce dernier a fait l'objet d'une cession de droit. C'est dans ce contexte de faible application de la LRN⁸ que nous avons lancé un chantier d'imports des publications récentes. Il consiste d'abord à verser dans HAL les métadonnées des publications et ensuite à contacter les auteurs en les incitant, conformément à la LRN, à déposer le texte intégral. Des chantiers d'import dans HAL et de quête du texte intégral ont déjà été réalisés par d'autres établissements, relevons l'Ifremer, l'École des Ponts⁹ et l'Université Rennes 1, laquelle met à disposition deux outils clés, OverHAL et Zip2Hal¹⁰, que nous avons largement utilisés avant et pendant le développement de notre solution.

LE CONTEXTE À L'UVSQ

L'UVSQ, qui rejoindra l'université de Paris-Saclay en 2025, développe une recherche pluridisciplinaire avec 35 structures de recherche, dont 13 unités mixtes de recherche (UMR), et 1 000 chercheurs ou enseignants-chercheurs. En 2015, le SCD de l'université crée le portail HAL UVSQ¹¹ dans lequel sont valorisés les laboratoires et projets de recherche portés par l'université à l'aide de collections dédiées.

Avant ce chantier d'import des publications récentes, le SCD réalisait des imports rétrospectifs des UMR. Le premier fut réalisé à partir d'une liste bibliographique BibTeX fournie par un laboratoire de mathématique : après un enrichissement manuel, les données étaient importées avec l'outil X2HAL de l'INRIA¹². Ensuite, les traitements se sont progressivement automatisés en utilisant Scopus comme

5 Une des forces de HAL tient dans son API, une technologie permettant d'accéder simplement aux données structurées de l'archive. En ligne : <https://api.archives-ouvertes.fr/docs/search>.

6 Afin d'être comparable avec la LRN, nous filtrons sur les articles scientifiques. Le pourcentage obtenu résulte du nombre d'articles déposés en texte intégral par les auteurs divisé par le nombre total d'articles dans HAL. Pour le numérateur voir la requête : https://api.archives-ouvertes.fr/search/?rows=0&fq=docType_s:ART&fq=submitType_s:file&fq=selfArchiving_bool:true. Pour le dénominateur, il suffit de retirer les deux derniers filtres de la requête. Ajoutons ici que la quasi-totalité (97 %) des publications traitées dans ce chantier est des articles scientifiques – seuls 17 communications dans des congrès et 5 chapitres d'ouvrages ont été trouvés.

7 Pour une explication complète de la loi, voir le guide d'application réalisé par le comité pour la science ouverte. En ligne : <https://www.ouvrirlascience.fr/guide-application-loi-republique-numerique-article-30-ecrits-scientifiques-version-courte/> (consulté le 10 mars 2021).

8 Sur l'application de la LRN, voir également le rapport *L'édition scientifique de revues : plan de soutien et évaluation des effets de la loi du 7 octobre 2016*. En ligne : <https://www.ouvrirlascience.fr/ledition-scientifique-de-revues-plan-de-soutien-et-evaluation-des-effets-de-la-loi-du-7-octobre-2016/> (consulté le 24 février 2021), dans lequel on peut lire : « En conclusion, s'il est certain que la loi a établi de manière salubre des règles précises sur le dépôt d'articles par les auteurs dans les archives ouvertes et a clarifié sur ce point les rapports entre les éditeurs et les auteurs, elle est loin d'avoir eu un effet incitatif. »

9 Frédérique BORDIGNON, « Comment booster le taux de full-texts dans HAL... », *Carnet IST*, 2 janvier 2017. En ligne : <https://carnetist.hypotheses.org/947> (consulté le 11 mars 2021).

10 En ligne : <https://halur1.univ-rennes1.fr/index.php> (consulté le 10 mars 2021).

11 En ligne : <https://hal.uvsq.fr/>

12 En ligne : <https://x2hal.inria.fr/homepage>

source des publications et l'outil OverHAL pour produire la TEI (pour Text Encoding Initiative) HAL¹³. Une campagne d'envoi de courriels à destination d'un laboratoire en sciences humaines avait été réalisée pour inciter au dépôt du texte intégral. Après l'import des publications de trois laboratoires, une limite décisive s'est fait sentir sur la quantité de publications qu'il était possible de traiter. L'efficacité de ces imports rétrospectifs a été évalué en comparant le nombre de publications des UMR entre HAL et Scopus. En une année, 10 % des UMR avait été traitée alors que des lacunes avaient été repérées dans 30 % d'entre elles. En conséquence, la durée nécessaire pour traiter les données lacunaires des unités aurait-elle même été la source de nouvelles publications, si bien qu'avec ces imports rétrospectifs nous aurions in fine couru après le temps. Les imports rétrospectifs ont ainsi été remplacés par l'import des publications récentes. L'intérêt pour ces dernières est issu de deux prédicats : plus la publication est récente plus les auteurs sont intéressés par les questions de diffusion, visibilité, et plus une publication est récente moins il est difficile d'obtenir une version antérieure à la version publiée.

Du point de vue technique, les métadonnées des publications récentes sont extraites de Scopus et automatiquement versées dans HAL. Les auteurs sont ensuite contactés afin qu'ils ajoutent le texte intégral de leur publication dans l'archive ouverte. La plupart des traitements sont effectués automatiquement à l'aide d'un code *ad hoc* réalisé avec le langage Python.

Ce code a été développé afin que d'autres établissements puissent l'utiliser : des connaissances rudimentaires en Python suffisent pour l'exécuter et il est mis à disposition sur la plateforme Github, laquelle favorise le développement collaboratif¹⁴. Le chantier d'import des publications a démarré en début de crise sanitaire, pendant le premier confinement : le travail limité en présentiel a permis à l'équipe « HAL UVSQ » d'augmenter ses forces.

Le chantier présenté répond à deux objectifs : d'une part, développer l'accès ouvert au sein de l'université et d'autre part augmenter la visibilité des publications, unités et projets de recherche émanant de l'université, grâce au portail HAL institutionnel.

ENJEUX TECHNIQUES : LA CHAÎNE DE TRAITEMENT

La chaîne de traitement se compose de trois étapes :

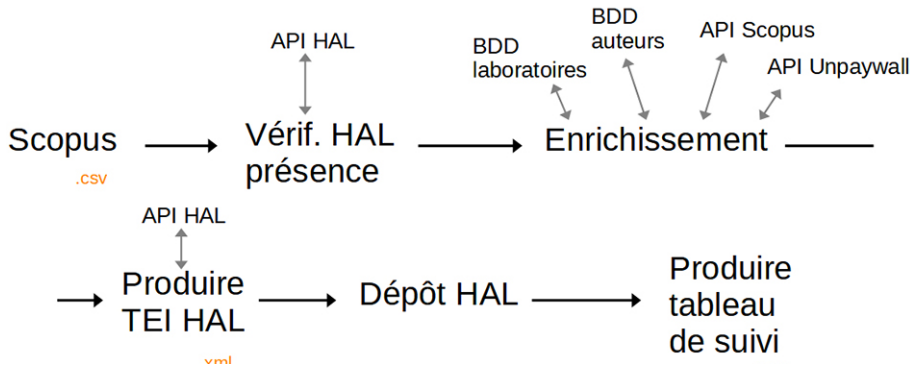
- 1 Récupération des métadonnées depuis Scopus, enrichissement et versement dans HAL ;
- 2 Complétion des métadonnées et ajout manuel du texte intégral ;
- 3 Envoi des courriels aux auteurs.

¹³ Documentation : <https://api.archives-ouvertes.fr/documents/all.xml>

¹⁴ Avec notamment le logiciel de gestion de version Git, le suivi de bugs ou encore la demande de fonctionnalités.

La première étape est celle qui nécessite le plus de technologie : elle consiste à récupérer les métadonnées de Scopus, vérifier la présence dans HAL, enrichir les métadonnées, effectuer le dépôt et finalement produire un tableau de suivi.

■ **Figure 2. Décomposition de la première étape de la chaîne de traitement**



Propre au paysage académique français, l'ajout des affiliations est un sujet majeur qui constitue une limite importante à l'automatisation des imports dans HAL. Par conséquent, un des objectifs principaux dans le développement du code a été de trouver une solution propre au périmètre de l'université, lequel regroupe toutes les UMR dont l'université est tutelle ou cotutelle. L'enjeu consiste à trouver automatiquement, à partir d'une affiliation textuelle brute, une structure du référentiel AureHAL¹⁵. Par exemple, l'affiliation suivante «Dynamiques patrimoniales et culturelles (DYPAC), Université Paris-Saclay, Versailles, 78000, France» doit être associée à la structure HAL ayant pour sigle et identifiant «DYPAC 418647». La solution réalisée recherche dans les signatures des éléments propres aux UMR du périmètre. Afin d'éviter les erreurs, nous avons opté pour une recherche à deux niveaux : la correspondance a lieu si des éléments propres à l'unité et aux tutelles sont présents. Le tableau suivant précise ces éléments pour chacun des niveaux.

Niveau de recherche	Éléments à rechercher
Unité de recherche	Forme longue, sigle, code
Tutelle	Forme longue, sigle

La recherche en forme longue s'effectue en recherchant plusieurs termes clés, et non par une recherche exacte. Par exemple, pour le DYPAC, la forme longue est valide si les trois termes «dynamique, patrimoniale, culture» sont présents. Enfin, si aucune affiliation n'a été trouvée, une recherche au niveau de l'université par sigle et forme longue est effectuée.

15 <https://aurehal.archives-ouvertes.fr/structure/index>

Le code fonctionne ainsi avec une table de données, à réaliser manuellement, qui décrit toutes les UMR¹⁶ du périmètre et permet d'effectuer les équivalences avec le référentiel AureHAL¹⁷.

La deuxième étape de la chaîne est réalisée manuellement. Elle se subdivise en deux sous-étapes : complétion et vérification des métadonnées HAL – par exemple, ajout des affiliations des auteurs extérieurs au périmètre de l'université – et ajout du texte intégral quand cela est possible ou bien prélèvement des courriels des auteurs à contacter. Comme il s'agit de traitements manuels, c'est l'étape la plus chronophage. Elle est réalisée en équipe, à l'aide du tableau de suivi produit à l'issue de la première étape.

La dernière étape consiste à contacter les auteurs dont les publications ne sont pas en accès ouvert et pour lesquelles la LRN est applicable. Les courriels contiennent le lien vers la notice HAL et expliquent l'applicabilité de la loi.

■ Figure 3. Un exemple de courriel envoyé

Bonjour,

Je travaille au développement de la science ouverte à l'Université de Versailles St-Quentin-en-Yvelines et notre équipe a récemment identifié un article dont vous êtes auteur :

- Bullying victimization and obesogenic behaviour among adolescents aged 12 to 15 years from 54 low- and middle-income countries
<https://hal.archives-ouvertes.fr/hal-03127505>

Cet article peut être diffusé en accès ouvert sur l'archive ouverte HAL : la loi "pour une République numérique" vous autorise en effet à partager 6 mois après publication [1] le manuscrit auteur accepté pour publication, c'est-à-dire la version qui précède la mise en forme éditeur [2].

Nous vous invitons à déposer dans HAL votre publication dès maintenant en appliquant un embargo. Pour cela veuillez nous communiquer l'email de votre compte HAL, afin que l'on vous accorde des droits nécessaires au dépôt, ou bien nous transmettre directement le fichier de votre manuscrit afin que nous le déposions pour vous.

Je reste à votre disposition pour toutes questions.

Bien cordialement,

Expéditeur

Équipe HAL UVSQ

DBIST

[1] Le délai varie selon le domaine : 6 mois après première publication pour le domaine des Sciences, de la Technique et de la Médecine et 12 pour les Sciences Humaines et Sociales. Sur ce sujet voir la vidéo de l'Université Paris-Saclay "Archives ouvertes : quid de mes droits d'auteur ?" <https://youtu.be/sv-NhB7tavs>

[2] Différences entre un manuscrit auteur accepté et la version éditeur : <https://hal.uvsq.fr/public/postprint.pdf>

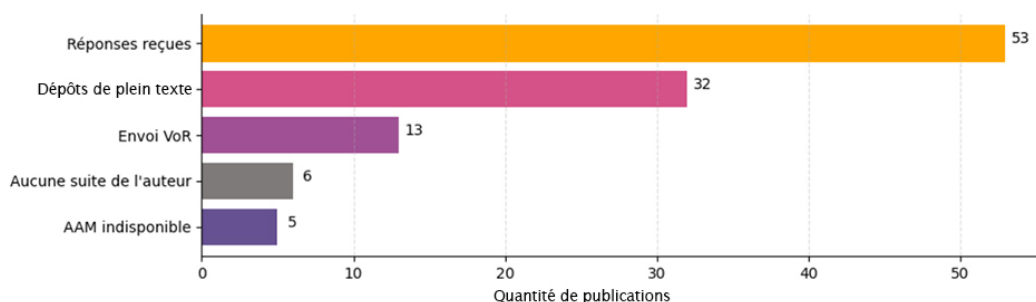
L'envoi des courriels est automatique, mais le suivi pour chacune des réponses obtenues est à réaliser manuellement. 352 courriels ont été envoyés en 10 mois. Nous avons eu un taux de réponse satisfaisant de l'ordre de 15 % et un taux de réalisation

16 En ligne: https://github.com/ml4rrieu/HAL_imports/blob/main/data/stable/labCriteria.csv (consulté le 2 mars 2021).

17 Ainsi que le référentiel des structures de Scopus si l'établissement possède une clé pour l'API.

– *i.e.* le nombre de PDF déposés par nous ou bien par l'auteur suite à notre sollicitation – de l'ordre de 9 %.

■ **Figure 4. Statistiques des réponses obtenues**



Légende : VoR = Version of Record et AAM = Author Accepted Manuscript.

La figure 4 permet d'identifier deux principales limites au dépôt du plein texte : l'absence de suivi de l'auteur et l'indisponibilité de la version auteur acceptée pour publication. La première limite renvoie à une plus globale : l'implication limitée des chercheurs ; la seconde nous montre une limite de la LRN : elle s'applique sur une version de publication que les chercheurs n'ont pas encore majoritairement intégrée (59 % des fichiers envoyés en première réponse par les auteurs correspondent à la version éditeur).

POLITIQUE D'IMPORT DANS HAL : QUAND LE CENTRE DE GRAVITÉ DU TRAVAIL BIBLIOTHÉCONOMIQUE SE DÉPLACE

Parmi l'ensemble des choix effectués pour importer dans HAL, deux s'avèrent non triviaux. Le premier concerne les affiliations des auteurs, le second l'ajout du texte intégral quand cela est possible.

Comme nous l'avons vu, la solution réalisée pour ajouter les affiliations fonctionne sur le périmètre de l'université ; les affiliations des auteurs extérieurs restent à compléter manuellement. En raison de l'hétérogénéité des affiliations françaises, cette opération est de loin la plus chronophage. Le problème est déjà bien connu¹⁸, il a pour cause première la structure complexe de l'enseignement supérieur et la recherche. Trouver une structure AureHAL correspondante à une affiliation renseignée par un auteur demande un temps non négligeable. De plus, en fonction des disciplines, certaines publications peuvent comporter plus d'une cinquantaine d'auteurs¹⁹. Après un mois de chantier, nous avons constaté que la complétion inté-

18 Voir les résultats du travail mené par les réseaux Renatis et l'URFIST de Paris en 2013 : D. PONTILLE, A. SÉNÉ, V. PRÊTRE *et al.*, « Éloge de la complexité : la signature des chercheurs et le millefeuille de l'affiliation institutionnelle dans les processus d'évaluation de la recherche », *Ethics, Medicine and Public Health*, juillet 2016, vol. 2, n° 3, p. 456-465. En ligne : <https://doi.org/10.1016/j.jemep.2016.07.008>.

19 300 auteurs est le maximum que nous avons trouvé. Cette donnée dépend évidemment des domaines disciplinaires de l'université.

grale des affiliations ne nous permettait pas de traiter suffisamment de publications. Conformément à la politique de HAL, qui impose a minima l'affiliation d'un auteur, il a été décidé de compléter uniquement les affiliations des auteurs français. Ce choix a soulevé des questions de management car par expérience un agent a tendance à être satisfait quand la notice sur laquelle il travaille est complète. Pour changer ces habitudes, rappeler l'objectif du chantier a été nécessaire : l'objectif du chantier n'étant pas d'obtenir des notices exhaustives, mais bien d'utiliser HAL pour développer l'accès ouvert.

L'ajout du texte intégral étant manuel, il a fallu définir des priorités : faut-il ajouter le fichier PDF si la publication est déjà en accès ouvert chez l'éditeur ? Que faire si elle est disponible dans une autre archive ouverte ? Nos choix ont été faits en suivant un principe de pérennité : seuls les publications en accès ouvert disponibles chez l'éditeur qui ne sont pas déjà présentes dans une archive sont à ajouter. En effet, les archives ouvertes sont mieux armées pour faire face aux épreuves du temps que ne peuvent l'être certaines revues scientifiques²⁰. Notre argumentaire a donc été d'assurer la pérennité des publications sans pour autant surcharger les traitements à effectuer²¹. Précisons enfin que seules les publications en accès ouvert avec une licence ouverte type Creative Commons sont ajoutées dans HAL : les publications en accès gratuit via le site de l'éditeur – correspondant au « bronze » dans la terminologie de Unpaywall²² – sont ainsi considérées comme « fermées » et un message d'incitation au dépôt est envoyé aux auteurs.

TRAVAIL COLLABORATIF ET CONFINEMENT

Avec plus d'une centaine de publications à traiter par mois, le chantier repose nécessairement sur un travail en équipe. Au sein du SCD, le contexte était favorable puisqu'à l'automne 2019 l'organigramme a été remanié en faveur du service à la recherche. L'équipe initiale « HAL UVSQ », qui réalise des missions variées, se composait alors d'un conservateur (responsable du service), de deux ingénieurs d'études (coadministrateurs) et deux personnels de catégorie B (responsables qualité des métadonnées). Mais le véritable catalyseur fut le premier confinement de mars 2020, conséquence de la crise sanitaire. Le confinement, limitant le travail en présentiel, a été l'occasion d'élargir l'équipe et de déployer le chantier : deux collègues de catégorie C, intervenant essentiellement sur la complétion et la qualité des métadonnées, ont rejoint l'équipe. Depuis le début du chantier, six agents sont ainsi mobilisés pour un équivalent temps plein d'environ 0,8²³. Une formation avancée a été nécessaire, non seulement pour

20 Sur la pérennité des revues en accès ouvert, voir Mikael LAAKSO, Lisa MATTHIAS et Najko JAHN, « Open is not forever : A study of vanished open access journals », *Journal of the Association for Information Science and Technology*, 2021. En ligne : <https://doi.org/10.1002/asi.24460>.

21 Les services de HAL offerts avec le PDF (table des matières, ajout des métadonnées de citation) sont aussi à prendre en compte dans cette direction.

22 Voir Heather PIWOWAR, Jason PRIEM, Vincent LARIVIÈRE *et al.*, « The state of OA », *PeerJ*, 6, 2018. En ligne : <https://doi.org/10.7717/peerj.4375>.

23 Avec la répartition suivante A et B : 0,2 et C : 0,4.

bien assimiler les processus techniques, mais aussi pour comprendre les enjeux de la science ouverte ainsi que les spécificités juridiques. Accompagner les collègues dans ces nouvelles fonctions a nécessité de faire preuve de réactivité et de disponibilité. En plein confinement, pour faciliter la communication, un groupe « WhatsApp » a été créé au début du chantier.

Le travail collaboratif s'organise autour de trois outils :

- le tableau de suivi des publications à traiter : un tableau partagé sur lequel chaque agent précise l'état de traitement de la publication ;
- une documentation des traitements à effectuer : un document qui rappelle les procédures à effectuer selon les informations des publications ;
- une adresse courriel en alias, permettant d'échanger entre collègues mais aussi d'assurer le suivi des courriels en équipe.

Ce projet demande une importante implication de l'équipe et la pérennisation d'un minimum de moyens humains pour ancrer la démarche sur le long terme. En septembre 2020, la contribution au chantier de collègues de catégorie C est acté dans les fiches de postes. La mise en place du chantier a permis une montée en compétences de tous les agents impliqués.

UN PANEL VARIÉ DE COMPÉTENCES

Les compétences nécessaires pour la réalisation de ce chantier sont essentiellement techniques et juridiques. De prime abord, il y a le substrat technique, le code Python, qui doit être exécuté périodiquement pour importer les métadonnées des publications récentes et envoyer les courriels. Après 8 mois d'utilisation, le code informatique conçu par le développeur a été revu afin d'être exécutable sans compétences en programmation. L'enjeu de cette révision était d'assurer la pérennité du service et de permettre à la communauté de s'approprier le chantier. Les compétences en Python pour exécuter le code sont donc rudimentaires : ajout de librairies, exécution de code, modification de variable textuelle pour préciser l'étape à exécuter²⁴.

Des compétences en traitement de données sont également à acquérir : « csv, UTF-8, json, xml, TEI » sont des éléments à maîtriser. Par exemple l'agent qui exécute le code peut être amené à modifier directement une information dans la TEI générée. Un éditeur de texte avancé pour lancer du code mais aussi lire et éditer des données formatées en JSON et XML est nécessaire.

Les autres compétences se situent à la croisée de la technique et du juridique et concernent l'écosystème des publications à l'heure de la science ouverte. D'une part, les éléments clés de cet écosystème doivent être acquis : DOI, PubMedId, ORCID,

24 Les ressources sur Python foisonnent, citons par exemple ce MOOC sur la plateforme FUN : « Apprendre à coder avec Python ». En ligne : <https://www.fun-mooc.fr/fr/cours/apprendre-a-coder-avec-python/> (consulté le 9 septembre 2021) ou encore le récent ouvrage d'Émilien SCHULTZ et Matthias BUSSONNIER, *Python pour les SHS : introduction à la programmation pour le traitement de données*, Presses universitaires de Rennes, 2021. En ligne : <http://www.pur-editions.fr/detail.php?idOuv=5092> (consulté le 23 février 2021).

ScopusId, les identifiants des projets de recherche, les versions d'une publication, la terminologie de SherpaRomeo, les licences Creative Commons ; et d'autre part, les aspects juridiques de l'accès ouvert doivent être maîtrisés. Par exemple, pour savoir si un texte intégral est à ajouter, l'agent doit différencier les informations suivantes, extraites de Unpaywall : « closed », « open from publisher no licence », « open in repository », « open from publisher : cc-by », « open from publisher : cc-by-nc » et « open from publisher : cc-by-nc-nd ». Par expérience, la différence entre une publication en accès gratuit via le site de l'éditeur et une publication disposant d'une licence Creative Commons n'est pas évidente pour les néophytes²⁵. De surcroît, la possibilité de déposer la version éditeur d'une publication qui n'est pas en accès ouvert sur le site de l'éditeur demande une bonne connaissance du sujet²⁶.

LIMITES ET AVANTAGES DU CHANTIER RÉALISÉ

Une couverture des publications perfectible

Une limite du chantier vient de l'utilisation de Scopus comme source de données : tout travail, qui utilise Scopus ou ses succédanés comme source, possède des limites de couverture. Scopus est en effet surtout adapté aux publications en sciences, techniques et médecine, avec une survalorisation des revues anglo-saxonnes²⁷. Ou mieux, utiliser Scopus c'est travailler avec les publications des éditeurs de courte traîne, c'est-à-dire les principaux acteurs de l'édition scientifique mondiale. C'est un problème conséquent pour lequel nous n'avons pas de solution aujourd'hui. Des services comme CrossRef²⁸, Conditor, ou les données du baromètre français de la science ouverte²⁹ pourraient dans un futur proche être utiles.

Un service et des professionnels de l'information plus visibles

L'avantage le plus important concerne le développement de l'accès ouvert. Cela se retrouve de façon évidente dans l'augmentation du nombre de publications déposées en texte intégral. Plus profondément, il faut surtout considérer les échanges que nous avons eus avec les chercheurs, les directeurs d'unités, les collègues et personnels de recherche : ces échanges ont contribué à une meilleure connaissance de l'accès ouvert, de ses enjeux, des aspects techniques et juridiques, ce qui est favorable au développement de l'accès ouvert.

25 À cet égard, l'archive ouverte HAL, du moins la version actuelle (2021-02), où le logo d'open access est ajouté pour ces deux types de notices, ne facilite pas la distinction.

26 Ce qu'autorise par exemple le *Journal of Geophysical Research : Atmospheres*. En ligne : <https://aurehal.archives-ouvertes.fr/journal/read/id/102531>.

27 Le baromètre science ouverte de l'UVSQ nous a permis de relever que Scopus couvrait, au maximum, 66 % de la production scientifique de l'université. En ligne : https://github.com/ml4rrieu/barometre_science_ouverte_uvsq (consulté le 23 février 2021).

28 Actuellement, CrossRef possède des lacunes trop importantes sur les données d'affiliations.

29 Voir Eric JEANGIRARD, *Monitoring Open Access at a national level : French case study, ELPUB 2019 23d International Conference on Electronic Publishing*, OpenEdition Press, 2019. En ligne : <https://hal.archives-ouvertes.fr/hal-02141819> (consulté le 11 mars 2021).

Un deuxième avantage concerne la visibilité des publications importées. Avoir à disposition des métadonnées de publications de son établissement dans HAL est utile à plusieurs égards : d'abord ces métadonnées sont enrichies des référentiels auteurs³⁰, structures et projets de recherche, ce qui permet d'obtenir aisément des listes de publication (*e.g.* rapports HCÉRES, compte rendu de chercheur, réutilisation dans le site web de l'université³¹ ou des laboratoires). Ensuite, ces métadonnées permettent d'éclairer les pratiques de publication existantes au sein de son établissement. Enfin, nous pensons aux systèmes d'information recherche qui peuvent utiliser en entrée les métadonnées de HAL³².

Le dernier avantage concerne la visibilité, non pas des publications, mais bien du service à la recherche. Un des enjeux de ce dernier est de se faire identifier par les chercheurs mais aussi plus largement par les différents acteurs de la recherche – UMR, service de la direction de la recherche, commission recherche, gouvernance. Contacter les auteurs quelques semaines après leurs publications renforce indubitablement la visibilité du service. Ainsi, le nombre d'interventions effectuées en laboratoire pour sensibiliser à l'accès ouvert a doublé, et les formations individuelles sur HAL ont également augmenté.

CONCLUSION

Après 10 mois de fonctionnement et bientôt 1 000 publications traitées, nous pouvons désormais revenir sur les objectifs fixés. Concernant le gain de visibilité, grâce à la structuration fine de HAL et notre travail sur les collections, cet objectif est pleinement atteint. Concernant l'accès ouvert, l'objectif principal pourrait paraître limité au regard du faible taux de réalisation : 9 % des courriels envoyés aboutissent à un dépôt effectif du texte intégral. Mais cet objectif ne peut se réduire à la présence d'un fichier PDF : il passe surtout par la sensibilisation à l'accès ouvert des différents acteurs de la recherche. De ce point de vue, les 15 % de réponses obtenues aux courriels envoyés, les demandes de formations individuelles et d'interventions en laboratoire, sont les signes d'un développement. Par ailleurs, grâce aux réponses obtenues, nous pouvons à présent ajuster nos échanges avec les chercheurs et doctorants : insister par exemple sur l'importance de la version acceptée pour publication, objet qui reste peu identifié par les chercheurs.

Concernant les perspectives, maintenant que l'équipe est stabilisée, les compétences acquises et les bénéfices démontrés, le SCD doit à présent communiquer

30 Au fil des imports, le code alimente une base de données (BDD) sur les auteurs affiliés à l'UVSQ. Après 10 mois de chantier, cette base intègre plus de 1 100 auteurs. Afin d'augmenter la visibilité de ces derniers, un alignement avec l'identifiant IdRef de l'ABES a récemment été effectué : 70 % des auteurs ont pu être alignés automatiquement.

31 Les métadonnées de HAL sont reprises dans le site général de l'université. En ligne : <https://www.uvsq.fr/laboratoire-atmospheres-milieus-observations-spatiales-latmos-umr-8190-cnrs-sorbonne-universite-uvsq> (consulté le 11 mars 2021).

32 Voir la récente journée d'étude « Vivo, une alternative aux solutions propriétaires pour diffuser librement les métadonnées de la recherche », avec les retours d'expérience de l'EHESS et l'UQAM. En ligne : <http://devlog.cnrs.fr/jdev2020/t5.p20210218> (consultée le 24 février 2021).

sur ce service d'import des publications récentes, le rendre plus visible au niveau politique. Une communication ciblée auprès des différents acteurs politiques de la recherche est en cours de réalisation – commission recherche, service de direction de la recherche et directeurs des unités. D'autre part, une communication en interne auprès des collègues du SCD est également à prévoir. Il est en effet nécessaire de sensibiliser l'ensemble des personnels du SCD à l'évolution de nos métiers et aux enjeux de l'accès ouvert. Enfin, à moyen terme, grâce à la visibilité gagnée, l'objectif est de définir une politique science ouverte à l'UVSQ, dont ce service d'import des publications récentes serait un des éléments principaux. •

Tous les contenus publiés sur le site du Bulletin des bibliothèques de France sont placés sous licence libre CC BY-NC-ND 2.0 : Attribution – Pas d'utilisation commerciale – Pas de modification 2.0 France.