



HAL
open science

HALUR : une suite d'outils connectés à HAL au service des chercheurs

Laurent Jonchère, Olivier Troccaz

► **To cite this version:**

Laurent Jonchère, Olivier Troccaz. HALUR : une suite d'outils connectés à HAL au service des chercheurs. JRES (Journées réseaux de l'enseignement et de la recherche) 2024, Renater, Dec 2024, Rennes, France. ⟨hal-04893970⟩

HAL Id: hal-04893970

<https://hal.science/hal-04893970v1>

Submitted on 17 Jan 2025

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 CC BY-NC 4.0 - Attribution - Non-commercial use - International License

HALUR : une suite d'outils connectés à HAL au service des chercheurs

Laurent Jonchère

Service Commun de la Documentation

Université de Rennes

4, rue Lesage CS 36402

35064 Rennes cedex

Olivier Troccaz

CREAAH UMR CNRS 6566

Avenue du Général Leclerc – CS 74205

35042 Rennes Cedex – France

Résumé

L'Université de Rennes, dans le cadre du Plan National de la Science Ouverte, vise un accès ouvert à 100 % de la production scientifique française d'ici 2030. Pour atteindre cet objectif, deux ingénieurs, Laurent Jonchère et Olivier Troccaz, ont développé HALUR, une suite d'outils PHP/Javascript utilisant les webservices de HAL. Ces outils facilitent le dépôt et la gestion des publications dans l'archive nationale HAL, permettant aux chercheurs de valoriser et rendre accessibles leurs travaux.

Les outils principaux sont OverHAL et Zip2HAL, qui alimentent massivement HAL via SWORD avec des métadonnées issues de bases de données internationales. CrossHAL enrichit les dépôts existants en utilisant les API de CrossRef et Pubmed. OverHAL envoie des messages personnalisés aux auteurs pour obtenir le texte intégral de leurs articles dans un format autorisé par les éditeurs, afin de les ajouter dans HAL.

MyHAL et ExtrHAL sont respectivement destinés aux chercheurs individuels et aux unités de recherche pour vérifier et extraire des listes de publications, tandis que VisuHAL fournit des indicateurs bibliométriques aux laboratoires.

HALUR a célébré ses dix ans en 2023 et a contribué à ce que 81 % des publications de 2022 de l'Université de Rennes soient en accès ouvert, soit 16 points au-dessus de la moyenne nationale. La suite est hébergée sur les serveurs de l'Université de Rennes et du CCSD/CNRS. Les scripts sont disponibles sur GitHub, bien que le dispositif ait des limites liées à sa maintenance et à l'absence de framework. Les concepteurs encouragent les propositions pour améliorer et pérenniser ces outils.

Mots-clefs

HAL, archive ouverte, accès ouvert, applications web

1 Les enjeux de l'accès ouvert

La science ouverte peut être définie comme la diffusion sans entrave des résultats, des méthodes et des produits de la recherche. L'Université de Rennes est pleinement engagée dans le mouvement de la science ouverte, pour atteindre l'objectif fixé par le Plan National de la Science ouverte (PNSO)

2021-2024)¹, mis en œuvre par le MESRI, d'un accès ouvert à 100% de la production française en 2030. Depuis la signature en 2013 d'un accord entre les universités, les grandes écoles et l'Académie des Sciences, tous les travaux scientifiques français ont désormais vocation à être référencés et diffusés dans l'archive nationale commune HAL². C'est la plateforme développée par le Centre pour la Communication Scientifique Directe (CCSD), sous l'autorité du CNRS, de l'INRAE et d'Inria.

Au sein de l'écosystème HAL, l'Université de Rennes a créé dès 2012 un portail institutionnel HAL-Rennes³ qui valorise la production des unités de recherche dont elle a la tutelle. L'objectif est double : créer un catalogue exhaustif de cette production, ce qu'aucune base de données actuelle ne permet, mais qui est essentiel pour une institution. On vise aussi à offrir un accès gratuit au contenu intégral de ces productions, en particulier pour les articles et les communications. La diffusion de la recherche est en effet depuis longtemps entravée par des coûts d'abonnements vertigineux, imposés par des éditeurs en situation de monopole. L'Université de Rennes dépense plus d'un million d'euros par an à ce titre. Par comparaison, l'éditeur Elsevier réalise chaque année un bénéfice net de 38% sur un chiffre d'affaires STM (Sciences Techniques et Médecine) de plus de 3 milliards d'euros... HAL apparaît donc comme une alternative vertueuse pour redonner à tous un accès libre et gratuit aux résultats d'une recherche financée en grande partie sur des fonds publics.

Son impact ne se cantonne pas au domaine français : la plateforme HAL est très efficacement indexée par les grands moteurs de recherche généralistes (Google) et spécialisés (Scholar, Pubmed, OpenAlex, etc.), et protégée contre les robots du web. La plupart des consultations et téléchargements sont effectués *via* ces moteurs, sans passer directement par HAL. Le portail HAL-Rennes bénéficie donc d'une visibilité à l'échelle mondiale. Par ailleurs, la valorisation ne se limite pas aux revues et aux communications dans un congrès, on y trouve aussi des chapitres d'ouvrage, des ouvrages, des thèses, etc. Depuis 2021, le deuxième PNSO promeut explicitement l'ouverture et le partage des codes sources et des logiciels.

Cette visibilité dépasse le périmètre des bouquets éditoriaux réservés aux seuls abonnés. Elle contribue ainsi au rayonnement de l'activité scientifique de l'université à l'international. Le dernier classement de Leiden⁴ place Rennes en sixième position en pourcentage de publications en accès ouvert pour la France (édition 2024).

L'existence d'un portail HAL institutionnel offre une autre opportunité : inventer de nouveaux services aux chercheurs, comme l'extraction de listes bibliographiques, le dépôt de publications par lots, la génération de tableaux et graphes statistiques et l'ajout de fichiers à des notices HAL sans texte intégral.

2 La suite HALUR

Pour répondre à ces défis, 2 ingénieurs, Laurent Jonchère (Service Commun de la Documentation - SCD / Université de Rennes) et Olivier Troccaz (UMR CNRS 6556 CREAAH) ont créé une suite d'outils PHP / Javascript baptisée HALUR⁵, basée sur les différentes API (Application Programming Interface) du CCSD⁶. Le point d'entrée de toutes ces applications est le code

1 <https://scienceouverte.univ-rennes.fr/le-cadre-national>

2 <https://hal.science/>

3 <https://univ-rennes.hal.science/>

4 <https://www.leidenranking.com/ranking/2024/list>

5 <https://ccsdrennesvip.ccsd.cnrs.fr/>

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

collection HAL (champ 'code_s' de l'API) correspondant à l'archive ouverte d'une unité, d'une équipe ou d'une institution. Ainsi, le portail HAL-Rennes a pour code UNIV-RENNES, et un code a été attribué à toutes les unités rattachées à l'université, aux équipes des unités, et aux unités d'appui et de recherche (UAR). Un même dépôt peut donc être accessible à travers différentes collections ou vues, selon un système de tamponnage qui appose automatiquement, sur la base d'une requête, un code collection. Il n'est donc heureusement pas nécessaire de refaire à chaque fois un dépôt pour chacune de ces différentes collections.

La suite HALUR s'est constituée au fil du temps, et n'a pris son nom que sur le tard. Elle désigne l'ensemble des outils créés successivement, au gré de l'inspiration de ses concepteurs, et des besoins internes ou en provenance des chercheurs. Laurent Jonchère, en tant qu'administrateur du portail HAL-Rennes, est chargé de la conception, et pour la partie technique, de la recherche des requêtes API ainsi que du test des applications « beta ». Olivier Troccaz, ingénieur de recherche en informatique, participe à la conception et écrit les programmes en PHP / JavaScript, dont il corrige les bugs par retours de tests. Pour plus de clarté, notre article fait le choix d'une présentation fonctionnelle de ces outils, en laissant de côté l'aspect chronologique de leur création. A noter que les 2 plus anciens, ExtrHAL et OverHAL, ont été construits sur des scripts élaborés par Philippe Gambette (Maître de conférences à l'université Gustave Eiffel)⁷, auquel nous sommes redevables. Ils ont cependant été transformés en profondeur de sorte qu'ils n'ont plus qu'un lointain rapport avec leurs modèles d'origine. La suite HALUR (UR pour « Université de Rennes ») compte aujourd'hui 6 applications, dont les noms soulignent la relation forte à l'archive ouverte : OverHAL, Zip2HAL, CrossHAL, VizuHAL, MyHAL et ExtrHAL. D'autres scripts d'importance moindre et cependant très utilisés complètent ce dispositif, mais n'ont pas encore reçu de nom spécifique. HALUR est librement accessible sur le web, mais certains programmes nécessitent une authentification via un compte CAS du CCSD (les mêmes identifiants et mots de passe que pour se connecter à HAL). Certaines opérations ne sont accessibles qu'à des profils utilisateurs disposant de privilèges attribués par le CCSD (niveaux administrateur, référent et gestionnaire). Quelques fonctionnalités sont réservées au périmètre rennais, via reconnaissance des adresses IP, pour des raisons techniques ou de sécurité.

3 Référencer la production des unités de recherche dans HAL

Une fois le portail HAL-Rennes mis en service, le plus urgent était de l'alimenter. En 2012, on était en effet encore très loin de l'exhaustivité bibliographique, et les dépôts en texte intégral étaient fortement lacunaires. Avant l'émergence des premiers outils HALUR, nous alimentions HAL grâce à l'application X2HAL⁸ créée par Laurence Farhi (Inria). En 2020, nous avons mis au point une chaîne complète pour alimenter HAL de manière autonome, en nous affranchissant de X2HAL : OverHAL et Zip2HAL. L'alimentation s'effectue en 2 temps. On dépose d'abord dans HAL les métadonnées de la publication (titre, auteurs, résumé, mots-clés, etc.), issues de bases de données bibliographiques internationales. Ces références sont ensuite complétées par le texte intégral, via une autre procédure sur laquelle nous reviendrons plus bas.

Pour la partie « métadonnées », l'émergence en 2022 d'OpenAlex⁹ (développée par l'organisme à but non lucratif OurResearch) a changé la donne : cette base de données bibliographique ouverte a pour ambition d'indexer la production scientifique mondiale de manière ouverte, transparente et réutilisable. Les données sont dans le domaine public CC0. Sa couverture est supérieure aux autres

7 <http://igm.univ-mlv.fr/~gambette/>

8 <http://doc.hal.science/en/x2hal/#>

9 <https://openalex.org/>

bases. Pour ces raisons, OpenAlex est devenu la source principale de nos dépôts HAL. Dans l'application OverHAL, l'utilisateur interroge directement l'API d'OpenAlex et récupère un fichier CSV des dernières publications de l'université ou d'un laboratoire spécifique. Le script analyse le fichier pour en extraire toutes les publications qui ne figurent pas encore dans la collection HAL, et les encapsuler dans un nouveau dossier *zip* XML-TEI conforme au modèle XSD du CCSD. Le dossier *zip* contient autant de fichiers TEI qu'il y a de publications. En complément, l'utilisateur peut également charger dans OverHAL un CSV ou un TXT extraits d'autres bases de données, comme Pubmed (base de référence en accès ouvert pour les sciences du vivant) et Web of Science (WoS), outil propriétaire auquel l'université est abonnée. L'utilisation conjointe de ces 3 sources assure une bonne exhaustivité du référencement, même si cela est beaucoup moins vrai pour les sciences humaines et sociales, et en particulier pour le droit. Dans ces disciplines, il n'existe pas vraiment de bases bibliographiques exploitables.

L'utilisateur enregistre le dossier *zip* dans l'outil Zip2HAL pour contrôler et enrichir chacun de ces TEI, via une interface dédiée, avant de les envoyer dans HAL. Sur une seule page web, le script affiche successivement les métadonnées de chaque item, dans un formulaire éditable à 6 colonnes. On y trouve le type de document (par exemple, un article de revue), les métadonnées (titre, année, résumé, etc.), le lien pérenne vers l'article sur le site de l'éditeur (son DOI ou Digital Object Identifier), les auteurs et leurs affiliations, le lien vers le TEI enregistré en local, et enfin un bouton « action » qui permet d'envoyer le TEI vers HAL. Ces tableaux sont précédés d'une liste de requêtes API sous forme de liens, qu'on peut consulter pour voir comment fonctionne l'algorithme de l'application. Zip2HAL calcule par essais successifs les identifiants des auteurs et des structures (affiliations), en cherchant des correspondances entre les données sources en *input* (noms des auteurs, leurs identifiants éventuels et les affiliations) et les données cibles (identifiants auteurs et structures dans les référentiels HAL). Les données sources sont parfois des chaînes de caractères, et heureusement de plus en plus des identifiants internationaux normalisés (ORCID pour les auteurs, ROR pour les structures).

Les propositions sont reportées dans le formulaire, charge à l'utilisateur de les vérifier et le cas échéant de les corriger. Les affiliations « source » sont affichées en grisé sous les affiliations calculées par le script, ce qui permet de les comparer et de repérer les éventuelles erreurs. La présence quasi systématique d'identifiants ROR et ORCID dans les notices OpenAlex facilite considérablement le travail de saisie et de vérification. Elle limite les erreurs de calculs, ce qui est appréciable quand on a une trentaine d'auteurs pour un même article, chose assez fréquente dans certaines disciplines ! Elle permet aussi d'affilier tous les auteurs, ce qui est loin d'être courant dans HAL. Beaucoup d'auteurs se bornent en effet à indiquer leur propre affiliation, négligeant celles de leurs co-auteurs.

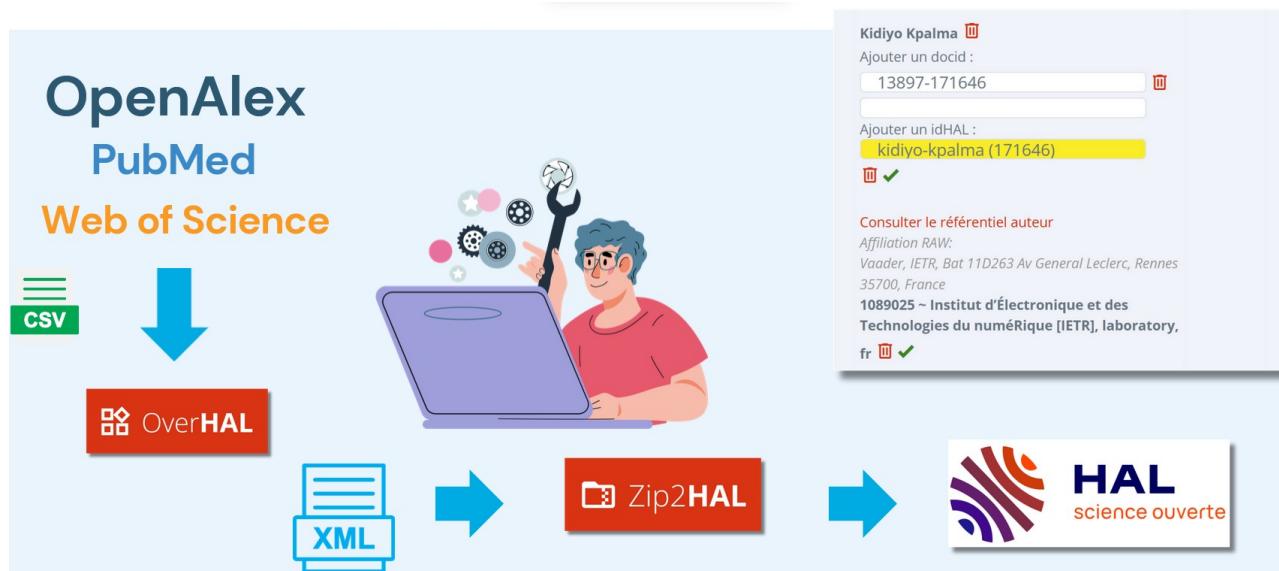


Figure 1 - Workflow OverHAL-Zip2HAL

Une fois la vérification effectuée, il ne reste plus qu'à cliquer sur le bouton « action » afin d'envoyer la publication dans HAL, via une requête HTTP POST dans l'API SWORD (Simple Web-service Offering Repository Deposit), au moyen d'une commande cURL. Dans certains cas, une erreur s'affiche en retour, signalant l'échec du dépôt, par exemple s'il manque un champ obligatoire. A ce stade, la majorité de ces dépôts sont donc des métadonnées sans texte intégral. OverHAL et Zip2HAL sont principalement destinés à un administrateur qui dépose dans HAL la production scientifique des unités de recherche, à la place des auteurs. Cela n'empêche pas les auteurs de déposer eux-mêmes leurs travaux, parfois *via* ces mêmes outils HALUR. A l'Université de Rennes, cette opération est effectuée une fois par mois, afin de limiter la taille des échantillons à traiter (limitée à 2 Mo pour les CSV et les dossiers *zip*) et suivre le rythme de la production institutionnelle.

4 Ajouter et diffuser le texte intégral des articles

La seconde étape, comme on l'a vu précédemment, consiste à récupérer et à ajouter le texte intégral, selon une méthode semi-automatisée. En France, la Loi pour une République numérique (article 30)¹⁰ autorise depuis 2016 la diffusion dans HAL du manuscrit auteur accepté (MAA) mais pas du PDF éditeur publié, sauf si celui-ci a été publié sous une licence ouverte Creative Commons (CC-BY)¹¹. Le MAA est une version relue par les pairs et acceptée pour publication, sans les corrections finales des épreuves (sauf si elles sont significatives, auquel cas nous les ajoutons). Problème : la communauté scientifique dans son ensemble ne se précipite pas pour déposer dans HAL ces MAA de façon massive, et en toute autonomie. Il faut donc un accompagnement, et souvent même déposer à la place des chercheurs.

Chaque semaine, l'administrateur HAL-Rennes fait ainsi une veille sur les plateformes éditeur afin de repérer les derniers articles publiés de son institution. Il les collecte dans l'outil de gestion bibliographique *open source* Zotero, en y ajoutant les adresses mail des auteurs à contacter. Il importe ensuite ces données CSV dans OverHAL pour envoyer des messages personnalisés aux

¹⁰ <https://scienceouverte.univ-rennes.fr/que-dit-la-loi>

¹¹ https://fr.wikipedia.org/wiki/Creative_Commons

auteurs de correspondance, les invitant à lui transmettre le MAA par retour de mail. En cas de non-réponse, un mail de relance est envoyé. Tous ces messages sont paramétrés dans des fichiers *template* PHP, et un code « labo Y » saisi dans Zotero permet de personnaliser chacun des mails par une phrase de type « Nous travaillons en collaboration avec Mme X, directrice de l'UMR Y », propre à mettre en confiance le destinataire. On peut même indiquer la langue dans laquelle est rédigé le message (français ou anglais), en fonction du domaine mail du destinataire et d'autres critères. Les personnes ciblées sont en effet des auteurs de correspondance d'articles scientifiques, qui peuvent être de nationalités très diverses, et bien sûr pas nécessairement francophones. Un fichier PHP liste les destinataires à exclure, heureusement en très petit nombre, qui ne souhaitent pas être recontactés pour diverses raisons. Plus important, OverHAL vérifie avant de lister les mails à envoyer, que lesdits articles n'ont pas déjà fait l'objet d'un dépôt en texte intégral dans HAL, ou d'un envoi précédent (contrôle sur le DOI). Côté utilisateur, l'interface OverHAL affiche une liste des articles concernés, avec un bouton « action » pour chaque entrée qui déclenche l'envoi du message.

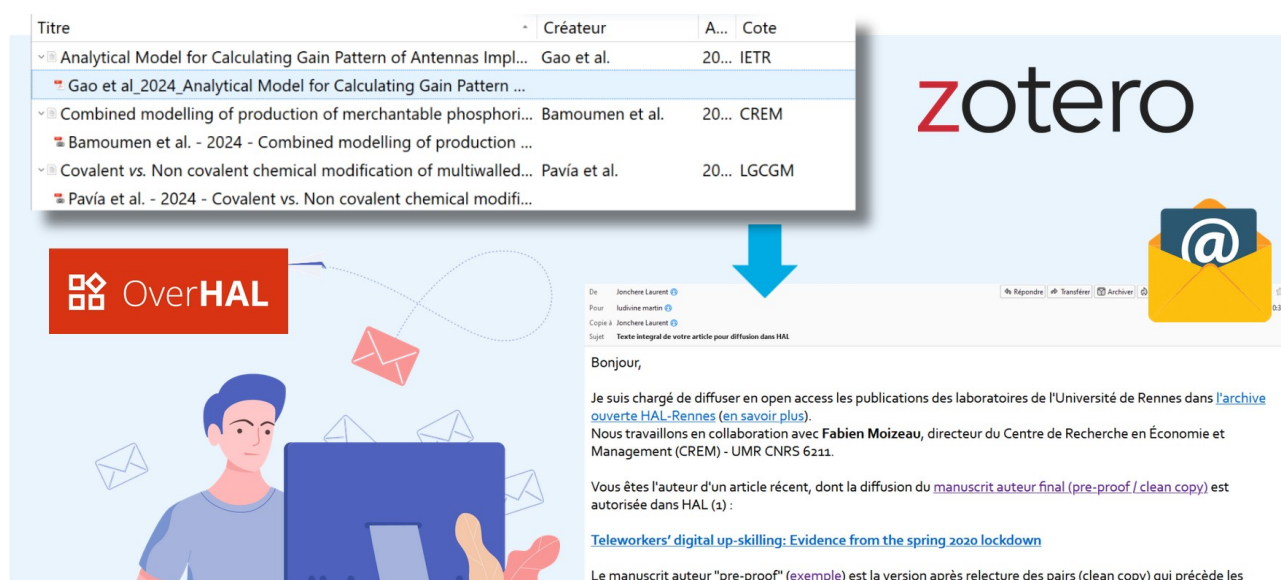


Figure 2 - Workflow Zotero - OverHAL

Cette procédure Zotero-OverHAL est un vrai succès, c'est même un des piliers sur lequel repose la stratégie d'ouverture des publications de l'Université de Rennes. Elle permet de récolter chaque année plus de 800 MAA, manuscrits qui sont ensuite ajoutés aux références bibliographiques HAL déposées dans la première procédure (OverHAL – Zip2HAL). Les auteurs sont généralement enthousiastes et approuvent la démarche¹². Autre avantage, les messages amenant souvent des questions de la part des intéressés, elle donne la possibilité d'engager avec eux des échanges fructueux sur la science ouverte, et d'en faire la pédagogie et la promotion.

5 Assurer la qualité des métadonnées

Une troisième brique multifonctionnelle, CrossHAL, a été initialement développée pour compléter, en aval, les dépôts des collections HAL, alimentées par nos outils ou par des tiers. Dans HAL, en effet, tous les champs de métadonnées ne sont pas obligatoires, mais certains sont importants

¹² <https://scienceouverte.univ-rennes.fr/centralisation-par-le-scd>

(résumé, mots-clés, financements, pagination finale, etc). Par ailleurs, on trouve aussi quelquefois des métadonnées erronées. Sans outil, il serait très fastidieux pour chaque dépôt de repérer ces champs, et encore plus de rechercher les données manquantes sur les sites des éditeurs pour les recopier ensuite dans les notices HAL. Fort heureusement, CrossHAL est capable d'effectuer automatiquement ces opérations d'enrichissement et de correction, en exploitant simultanément les API de CrossRef¹³ (agence à but non lucratif d'enregistrement et de registre des DOI) et de Pubmed, sur la base de requêtes sur les DOI HAL. En résultat, le script soumet à l'utilisateur une liste des publications HAL concernées dans la collection interrogée, sous la forme d'un tableau comparé des métadonnées HAL / CrossRef ou Pubmed. Par exemple, pour tel dépôt, la pagination en gras dans la colonne CrossRef indique que la métadonnée HAL correspondante est absente ou erronée. On peut choisir quels champs contrôler. L'outil est capable d'analyser de 10 à 200 notices à la fois, en tenant compte cependant des limitations imposées par CrossRef aux volumes d'interrogations simultanées. Un bouton action permet de valider la proposition d'un clic pour chaque dépôt, selon la méthode cURL PUT du protocole SWORD, mais l'utilisateur pressé (ou fainéant !) peut cliquer sur un autre bouton en bas de page qui valide en une seule fois toutes les propositions remontées dans la tranche analysée. Ce dispositif de correction et complétion automatique garantit la qualité et la richesse descriptive du référencement.

Traitement référence 28 sur 200





















ID	Liens			AP	1er auteur	Revue	Vol(n)pp		PMID		Année de publication	
	DOI	HAL	CR				HAL	CR	HAL	Pubmed	HAL	CR
1							()	9(3)1427-1440			2024	2024-09
4				AP			()		39136832	39136832	2024	2024-08-13
8							()227-237	()227-237			2024-03-02	2024-02-17
11							17(3)668-675	17(3)668-675	38740182	38740182	2024	2024-05

Figure 3 - tableau d'analyse de CrossHAL (détail)

6 Ajouter les PDF éditeur sous licence ouverte

Le même outil CrossHAL est utilisé pour repérer dans une collection HAL les publications pour lesquelles existe un fichier PDF sous licence Creative Commons CC-BY, à partir d'une interrogation de l'API Unpaywall¹⁴ qui indexe l'url de ces fichiers. Il s'agit généralement d'articles publiés dans des revues en accès ouvert, ou bien chez des éditeurs avec lesquels l'Université a signé un accord pour la diffusion en accès ouvert dans une revue sous abonnement. Pourquoi les recopier dans une archive ouverte s'ils sont disponibles chez l'éditeur ? Essentiellement pour des raisons de souveraineté car rien ne garantit l'accès à long terme de ces fichiers, tandis que les PDF

13 <https://github.com/CrossRef/rest-api-doc>

14 <https://unpaywall.org/products/api>

déposés dans HAL font l'objet d'un archivage sécurisé sur les serveurs du Centre Informatique National de l'Enseignement Supérieur (CINES). On veut pouvoir garantir un accès pérenne, quelle que soit l'évolution des formats et supports de lecture. Comme pour le contrôle des métadonnées, le script affiche un tableau des items avec un bouton « action » permettant d'ajouter le lien dans HAL, encore une fois selon la méthode cURL PUT du protocole SWORD. Malheureusement, l'URL des fichiers fournie par les éditeurs dans CrossRef, que récupère Unpaywall, n'est pas toujours exploitable pour l'API HAL. Elsevier, par exemple, ajoute dans cette URL des données de session utilisateur pour mieux pister son activité. Néanmoins, cette fonctionnalité de CrossHAL est très appréciable, ne serait-ce que pour repérer les dépôts HAL éligibles.

7 Évaluer le taux d'accès ouvert aux publications

Quelle est la pertinence de toutes ces procédures et comment évaluer leur portée ? Chaque année depuis 2021, le Baromètre de la science ouverte (BSO) du MESRI mesure la progression de l'accès ouvert. Les résultats sont excellents, puisque dans la dernière version du BSO, 81% des publications 2022 de l'Université de Rennes étaient en accès ouvert, soit 16 points au-dessus de la moyenne nationale¹⁵. L'année 2022 sert de référence en raison des barrières mobiles encore en vigueur pour l'année 2023, et pour faciliter les comparaisons nationales. Un succès qui témoigne de l'efficacité des procédures de référencement-dépôt mises en place avec les outils HALUR. En parallèle, l'outil VisuHAL permet également d'évaluer le taux d'accès ouvert aux publications. Il a été créé pour fournir aux unités de recherche des indicateurs bibliométriques, sous forme de tableaux et de graphiques, à partir d'une vingtaine de requêtes API HAL prédéfinies : progression des dépôts, du taux d'accès ouvert, répartition par éditeurs, financements, collaborations internationales, etc. La page de résultat de chaque requête contient un bloc de présentation, le tableau ou les graphes de résultats, et une documentation technique indiquant le type de requête utilisé. Des infobulles sont parfois positionnées dans les cellules pour aider à contextualiser les chiffres. Les données des tableaux sont exportables en format CSV. En plus de l'API HAL, l'application VisuHAL recourt à l'API Unpaywall pour certaines requêtes. A noter que Thierry Fournier, alors responsable du service de gestion centralisée des abonnements du SCD de l'Université de Rennes, a contribué à faire évoluer VisuHAL en proposant des idées de requêtes pour aider à l'acquisition de ressources documentaires scientifiques.

8 Vérifier et extraire des listes de publications HAL

Destiné à un usage individuel, MyHAL est un outil très simple d'utilisation : un auteur peut consulter la liste de ses publications HAL et repérer celles pour lesquelles il manque le texte intégral, qu'il peut ensuite facilement ajouter d'un clic de bouton (« ADD »). Comme pour la plupart des autres applications HALUR, les données affichées en résultat sont issues de l'API HAL, en intégrant tous les cas de figure possibles des éléments fournis en *input* (exemple : prénom nom, p nom, p. nom, noms composés avec ou sans tiret, espaces superflus, noms et prénoms inversés, etc.), et sont exportables en CSV. L'interface est en anglais pour permettre un usage élargi à la communauté des scientifiques. La fonctionnalité ADD est particulièrement appréciée des chercheurs CNRS qui ont l'obligation de déposer (souvent dans l'urgence) le texte intégral de leurs

15 <https://scienceouverte.univ-rennes.fr/barometre-science-ouverte-edition-2023>

articles dans HAL, afin de valider le bilan annuel de leur production dans l'application CRAC du CNRS.

MyHAL est également utile pour préparer en amont les bilans d'évaluation du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres), à un niveau individuel par chaque auteur avant de remonter au niveau des équipes. A ce niveau supérieur, la brique ExtrHAL propose aux unités et équipes de recherche d'afficher en HTML et d'extraire, dans des formats RTF, PDF, CSV ou BibTeX, des listes de publications à partir d'un code collection HAL, toujours à partir de requêtes API HAL. Les métadonnées JSON retournées sont réagencées dans un format bibliographique spécifique à chaque type de document et sous-catégorie de document (articles, articles à comité de lecture, communications, posters, logiciels, etc.). L'utilisateur peut personnaliser le format de sortie (gras, souligné, avec certains liens, etc.), selon les besoins spécifiques de chaque discipline. ExtrHAL a été conçu comme un outil modulaire pour générer des rapports bibliographiques sur mesure, en quelques clics. En 2023, dix ans après sa création, l'outil a été repris et adapté par le Hcéres pour lancer l'application monevaluation.hceres.fr¹⁶, bénéficiant de nos conseils et de notre expertise.

9 D'autres outils non recensés dans HALUR

HALUR propose d'autres scripts très utiles, mais moins visibles car réservés à certaines catégories d'utilisateurs. Le plus important, `Publis-HAL-SCD.php`, est appelé en *iframe* dans le code « source » des pages « publications » des sites web des laboratoires affiliés à l'université. Il permet d'alimenter automatiquement ces pages, à partir de l'API HAL, pour afficher en temps réel la liste des publications d'une unité, d'une équipe ou d'un chercheur¹⁷. Précieux gain de temps, puisqu'il n'est plus nécessaire pour ces laboratoires de compléter et de mettre à jour ces pages « à la main », selon une procédure lourde et laborieuse.

Un second script, `depot_init.php`¹⁸, est destiné aux administrateurs HAL pour construire une requête de tamponnage. C'est le filtre API qui alimente une collection HAL. La requête est générée à partir d'un CSV formaté en UTF-8 contenant en colonnes la liste des prénoms et noms des chercheurs d'une équipe. Cela permet de gérer facilement les collections HAL pour lesquelles le seul critère affiliation ne suffit pas, car il faut resserrer aux auteurs d'une équipe particulière au sein d'un laboratoire. Pour donner un exemple simple, on obtient ainsi une requête booléenne de type : `(authFullName_t:"Prénom Nom" OR authFullName_t:"P Nom" OR etc) AND (structId_i:000000 OR etc)`. Une telle requête serait trop longue à rédiger à la main et pourrait d'ailleurs contenir des erreurs, invalidant le tamponnage. Quand des changements interviennent au sein d'une équipe, suite à des départs ou des entrées de personnels, il suffit de mettre à jour la requête à partir d'une nouvelle liste d'auteurs en format CSV.

16 <https://monevaluation.hceres.fr/hal>

17 Un exemple de page pour l'UMR CNRS 6226 : <https://iscr.univ-rennes.fr/publications-0>

18 https://ccsdrennesvip.ccsd.cnrs.fr/depot_init.php

10 Pérenniser ces outils au service de la communauté

La suite HALUR a fêté ses 10 ans en 2023. Elle est hébergée sur un serveur de la DSI de l'Université de Rennes¹⁹, et tourne également depuis 2023 sur une VM du CCSD, sur la proposition généreuse de son responsable de la production, Bruno Marmol. Les scripts sont référencés dans un GitHub sous licence GPLv3. Malgré quelques bugs occasionnels, propres à tout logiciel, et d'ailleurs vite résolus, les outils s'avèrent remarquablement stables et robustes à l'utilisation. Ses concepteurs ont choisi d'ouvrir et de partager largement HALUR à la communauté des établissements et des chercheurs, qui n'ont pas manqué de se les approprier. Des présentations et ateliers sont régulièrement proposés pour former les utilisateurs. Malgré son succès, le dispositif rencontre aujourd'hui quelques limites. Le support et la maintenance reposent essentiellement sur 2 personnes, et le code n'est pas encapsulé dans un *framework*, ce qui en limite la reprise et l'enrichissement par d'autres développeurs. Toutes propositions sont donc les bienvenues pour contribuer à la pérennité de ces outils au service de la recherche !

19 <https://halur.univ-rennes.fr/>