

# HALUR: des outils connectés à HAL au service des chercheurs

HALUR<sup>1</sup> est une suite d'outils créée par deux ingénieurs, Laurent Jonchère (Université de Rennes) et Olivier Troccaz (UMR CNRS 6566), centrée sur l'accès ouvert aux publications des chercheurs. Ces six applications Web exploitent les possibilités de l'interface de programmation (API) de l'archive ouverte nationale HAL<sup>2</sup>, où sont référencées les publications scientifiques de l'Université de Rennes, dont une majorité sont disponibles en texte intégral.

## LAURENT JONCHÈRE

Ingénieur d'études au Service commun de la documentation – Université de Rennes et administrateur du portail HAL-Rennes

## OLIVIER TROCCAZ

Ingénieur de recherche UMR CNRS 6566 Centre de recherche en archéologie, archéosciences, histoire (CRéAAH)

1. <https://halur1.univ-rennes1.fr/index.php>

2. <https://hal.science/>

3. <https://appui-evaluation.hceres.fr/hal> : outil dérivé d'une application similaire de l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), elle-même directement inspirée d'ExtrHAL.

La première brique HALUR a été créée en 2013 sous le nom d'ExtrHAL. Il s'agissait de répondre à un besoin urgent des unités de recherche d'exporter la bibliographie complète de leurs productions pour l'évaluation du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres). Nous avons adapté et enrichi un script du chercheur Philippe Gambette (Université Gustave Eiffel), croisé à Lyon lors d'une formation HAL.

Le principe sera repris sur tous les outils ultérieurs de HALUR : l'utilisateur saisit le code collection HAL de son laboratoire, correspondant à l'ensemble des publications de cette équipe, puis renseigne quelques paramètres (dates, type de document, etc.) avant de valider sa requête. L'outil, que nous avons souhaité partager avec la communauté universitaire, a rencontré un vif succès et a été adopté par un nombre important d'établissements et de laboratoires.



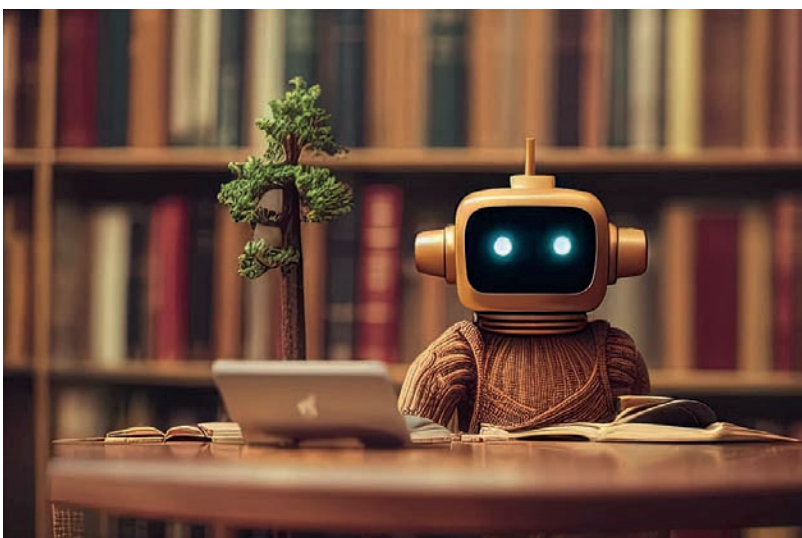
<https://unsplash.com> – Photo de Scott Graham

Nous avons été depuis lors très sollicités pour effectuer des mises à jour et apporter un support technique, mais aussi pour animer des formations. Dix ans plus tard, le Hcéres s'est inspiré d'ExtrHAL pour créer sa propre application d'export<sup>3</sup>, en s'appuyant sur notre expertise.

## Référencer la bibliographie des laboratoires dans HAL

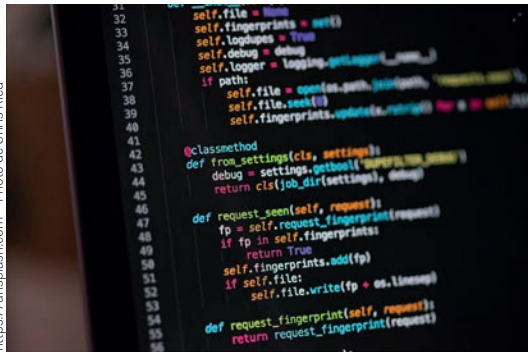
Encore faut-il, pour exporter la bibliographie d'un laboratoire ou d'une institution, que ses publications aient bien été toutes référencées dans HAL. C'était une gageure, car il n'était évidemment pas question de le faire article par article, et nous ne pouvions pas non plus compter sur une participation active de tous les chercheurs. Nous avons ainsi créé les deux briques d'une chaîne de dépôt massif, **OverHAL** et **Zip2HAL**, la première pour convertir en un format lisible par les machines (XML) des listes de publications issues de bases de données bibliographiques internationales (Web of Science, Pubmed), la seconde pour déposer en masse ces publications dans HAL *via* des webservice, après vérification des métadonnées.

Ces deux outils sont principalement destinés à un administrateur HAL de l'université qui se charge



<https://pixeid.com> – (licence CC-BY)

Un robot d'intelligence artificielle apprend dans une bibliothèque.



https://unsplash.com - Photo de Chris Ried

de déposer dans HAL la production scientifique pour le compte des unités, à la place des auteurs. L'administrateur HAL extrait ainsi chaque mois les nouvelles publications indexées dans les bases Web of Science et Pubmed, afin de les référencer dans HAL en utilisant la chaîne OverHAL-Zip2HAL. Un algorithme établit la correspondance entre des chaînes de caractères (noms d'auteurs, affiliations) et les personnes et laboratoires enregistrés dans les référentiels HAL. Grâce à ce dispositif, il nous a été possible d'atteindre assez vite l'exhaustivité du référencement bibliographique des unités de recherche, principalement sous forme de métadonnées (titre, auteur, résumé, etc.).

### Ajouter et diffuser le texte intégral des articles

Cependant, le but d'une archive ouverte est avant tout d'offrir un accès ouvert au texte intégral de ces publications, et pas de se limiter à un simple catalogue. Pour cela, nous avons fait évoluer **OverHAL** de manière à pouvoir collecter les manuscrits des articles auprès des auteurs. En effet, la loi pour une République numérique<sup>4</sup> autorise 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)<sup>5</sup>. Nous avons donc adopté une nouvelle procédure, qui complète celle du référencement de la chaîne OverHAL-Zip2HAL : l'administrateur HAL fait chaque semaine une veille sur les plateformes éditeur afin de repérer les derniers articles publiés, qu'il collecte dans l'outil libre Zotero<sup>6</sup>, en y ajoutant les adresses mail des auteurs à contacter. Il importe ensuite ces données dans OverHAL pour envoyer des messages personnalisés aux auteurs de correspondance, les invitant à lui transmettre le « Manuscrit auteur accepté » par retour de mail. En cas de non-réponse, un mail de relance est envoyé. Cette fonctionnalité, depuis sa mise en place en 2013, a permis de récolter un très grand nombre de « Manuscrits auteurs acceptés », qui ont pu être ajoutés aux références bibliographiques déjà enregistrées dans HAL. Elle a en outre offert l'avantage de mieux nous faire connaître auprès des auteurs, d'engager avec eux des échanges fructueux et de les informer sur les différentes possibilités de HAL.

### Assurer la qualité des métadonnées

En parallèle, une quatrième brique, **CrossHAL**, a été développée pour ajouter de façon systématique le PDF éditeur sous licence CC-BY, mais également pour enrichir les métadonnées des dépôts

existants d'une collection HAL, à partir de données de bases de données externes interrogeables avec des interfaces de programmation d'application (ou *Application Programming Interface* – API). Manière à la fois d'augmenter le taux d'accès ouvert des publications HAL et de veiller à la qualité et à la richesse descriptive du référencement.

### Évaluer le taux d'accès ouvert aux publications

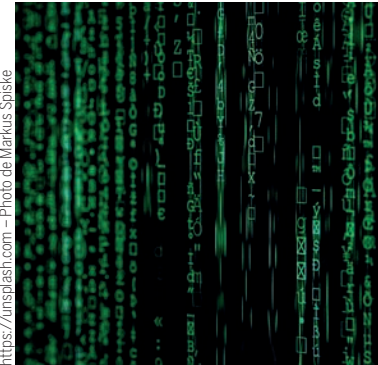
Toutes ces procédures ont été formalisées en 2017 dans une politique d'accès ouvert de l'Université de Rennes, invitant les chercheurs à ajouter ou faire ajouter le MAA de leurs articles dans HAL. Un nouvel outil, **VizuHAL**, a été développé à cette occasion<sup>7</sup> pour mesurer chaque année la progression de l'accès ouvert, avant que le Baromètre de la science ouverte (BSO)<sup>8</sup> du ministère de l'Enseignement supérieur et de la Recherche (MESR) n'en prenne le relais à partir de 2021 : dans la dernière édition du BSO, 80 % des publications 2021 de l'Université de Rennes étaient en accès ouvert, soit 13 points au-dessus de la moyenne nationale<sup>9</sup>. Succès qui témoigne de l'efficacité des procédures de référencement-dépôt mises en place avec les outils HALUR.

### Vérifier sa liste de publications HAL

La dernière brique, **MyHAL**, a vu le jour en 2020, suite à l'obligation faite aux chercheurs du CNRS de déposer le texte intégral de leurs articles dans HAL, afin de valider le bilan annuel de leur production dans l'application CRAC du CNRS. Il fallait leur offrir un outil simple d'utilisation, leur permettant de consulter d'un clic la totalité de leurs publications HAL et de repérer d'un coup d'œil les PDF déjà déposés. Comme la plupart des articles étaient déjà référencés dans HAL par nos soins, il ne leur restait plus qu'à ajouter le « Manuscrit auteur accepté » aux notices sans texte intégral, au moyen d'un petit bouton créé à cet effet. Plus largement, MyHAL est utilisé par tous les chercheurs de l'Université, affiliés ou non au CNRS.

### Pérenniser ces outils au service de la communauté

C'est ainsi qu'au fil des années, la suite HALUR s'est développée et enrichie, avec la volonté d'ouvrir ces outils à tous, autant que possible. À la base, nul cahier des charges, nul comité de pilotage ou autre *task force*. Juste la collaboration libre et informelle entre deux collègues passionnés, donnant forme à leur imagination à partir de besoins concrets. Revers de la médaille, la fragilité d'un dispositif qui repose essentiellement sur deux personnes, s'agissant de la maintenance et du support technique (mais l'aventure dure avec succès depuis 10 ans !), et sur une mémoire serveur limitée qui ne permet pas de traiter de gros volumes de requêtes ou de fichiers. Fort heureusement, des solutions émergent, ce qui devrait permettre à terme d'assurer leur pérennité et de mieux les partager avec la communauté des établissements représentés dans HAL<sup>10</sup>. ■



https://unsplash.com - Photo de Markus Spiske

4. <https://scienceouverte.univ-rennes.fr/que-dit-la-loi>
5. [https://fr.wikipedia.org/wiki/Creative\\_Commons](https://fr.wikipedia.org/wiki/Creative_Commons)
6. Zotero est un logiciel de gestion bibliographique complet, libre et gratuit.
7. Avec la contribution de Thierry Fournier, Directeur adjoint du service commun de documentation de l'Université de Rennes.
8. Voir dans ce numéro l'article de Laetitia Bracco, p. 26.
9. <https://scienceouverte.univ-rennes.fr/barometre-science-ouverte-edition-2022>
10. Une enquête Flash d'octobre 2021 de l'association CasuHAL regroupant les utilisateurs HAL plaçait ExtrHAL et OverHAL dans le top 7 des chantiers prioritaires pour HAL : <https://www.casuhal.org/category/enquetes/>