“ Avec HAL, nous voulions créer un arXiv multidisciplinaire ”

Daniel Charnay

▶ **To cite this version:**

Daniel Charnay. “ Avec HAL, nous voulions créer un arXiv multidisciplinaire ”. Hermès, La Revue - Cognition, communication, politique, 2019, La communication au cœur des connaissances, 2019/3 (85), pp.94-97. hal-02387841

HAL Id: hal-02387841
https://hal.science/hal-02387841

Submitted on 26 Sep 2023

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.

CNRS EDITIONS

« Avec HAL, nous voulions créer un arXiv multidisciplinaire »

Daniel Charnay

Dans Hermès, La Revue 2019/3 (n° 85), pages 94 à 97
Éditions CNRS Éditions

ISSN 0767-9513
ISBN 9782271129642
DOI 10.3917/herm.085.0094

Article disponible en ligne à l'adresse
https://www.cairn.info/revue-hermes-la-revue-2019-3-page-94.htm

CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour CNRS Éditions.
La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

© CNRS Éditions | Téléchargé le 24/09/2023 sur www.cairn.info (IP: 90.66.145.31)

Daniel Charnay

« Avec HAL, nous voulions créer un arXiv multidisciplinaire »

David Rochefort : Pouvez-vous revenir sur la naissance de HAL ?

Daniel Charnay : L'IN2P3 avait un système documentaire fourni par un prestataire centralisé à Lyon. En 1992, le contrat arrivait à terme et j'ai proposé de faire une application de documentation sur le Web. C'était les tout débuts du Web, on m'a certifié que ça ne marcherait jamais. Et pourtant, Démocrite – une application documentaire, qui ressemblait à une archive ouverte dans la mesure où il était possible de déposer du texte intégral – a été créé.

Mais celui qui a vraiment réussi à convaincre le CNRS de faire des archives ouvertes – et ce n'était pas gagné –, c'est Franck Laloé, chercheur à l'ENS. Il a convaincu Catherine Bréchignac, alors directrice générale du CNRS. Les premières discussions ont eu lieu en 1999.

Franck Laloé est un physicien, et il avait besoin d'informaticiens autour de lui. C'est par l'intermédiaire d'Édouard Brézin, ancien président du CNRS, que nous avons été mis en relation et que nous avons pu démarrer cette archive ouverte.

En 2000, nous avons ouvert HAL, avec des règles bien précises, largement transgressées par la suite : nous mettions à disposition une archive ouverte de textes intégral. Notre modèle, c'était arXiv, et nous voulions réaliser un arXiv multidisciplinaire.

Nous avons donc créé cette archive ouverte en 2000 et la première version a été mise à disposition en 2001 – d'où le nom HAL (Hyper Archives en ligne), qui fait référence au film de Kubrick.

David Rochefort : Est-ce qu'il a été facile de convaincre les chercheurs ?

Daniel Charnay : Nous avons créé cet outil et nous l'avons mis en ligne. C'était la première étape. Mais le travail le plus compliqué a effectivement été de convaincre. Pour le monde de la physique, ce fut assez facile. Les physiciens écrivent leurs articles dans un langage qui s'appelle LaTeX. C'est compliqué à manipuler ; quand ils déposaient leur texte, arXiv était assez rigoureux dans les vérifications. Si le texte était mal structuré, arXiv le renvoyait. Pour faire venir les physiciens, on leur a donc dit : « déposez du

© CNRS Éditions | Téléchargé le 24/09/2023 sur www.cairn.info (IP: 90.66.145.31)

LaTeX dans HAL, même s'il n'est pas bien conçu, on le réparera et on enverra la bonne version chez arXiv ». Dès le départ, nous avions eu l'idée de lier HAL à arXiv. Paul Ginsparg, son créateur, était venu nous voir, nous avions fait la connexion.

Comme les physiciens avaient cette culture du dépôt, des archives ouvertes, du preprint, il a donc été assez facile de les convaincre – même si certains nous disaient (et continuent) qu'on faisait un projet trop franco-français et trouvaient plus prestigieux d'aller directement chez arXiv.

Pour les autres disciplines, ça a été infiniment plus compliqué. Certaines ont été réticentes dès le début et le sont restées. Comme les chimistes. Il y a beaucoup de processus industriels en jeu, on doit d'abord valoriser, prendre des brevets.

David Rochefort : Et pour les sciences humaines ?

Daniel Charnay : Nous sommes allés assez vite, parce qu'il y avait un directeur adjoint scientifique assez favorable. HAL SHS a été ouvert et a relativement bien marché parce qu'en SHS il n'y avait rien. La difficulté, c'est que les SHS avaient tendance à déposer une grande variété de documents, pas toujours du niveau d'une publication scientifique. Le principe de HAL, c'est qu'un article devait être scientifique, avec des critères de scientificité qu'on avait établis. Y avait-il plusieurs auteurs ? Est-ce qu'ils émanaient d'institutions scientifiques reconnues ? Est-ce qu'il y avait une résumé, des références bibliographiques ? etc. On cherchait à savoir s'il avait la couleur d'un article scientifique, sans juger sur le fond – on en aurait été bien incapables. C'était plus compliqué en SHS parce que les chercheurs publient souvent seuls et sont souvent moins liés au niveau institutionnel. La modération a été plus difficile avec ce qui était déposé dans HAL SHS, moins automatisée.

David Rochefort : Sans faire d'évaluation ?

Daniel Charnay : En faisant quand même un peu d'évaluation, même si on ne le disait pas. Avec tous les problèmes propres aux sciences humaines : est-ce que les thèses d'un article sont révisionniste ou pas ? etc. Nous avions constitué un petit réseau de chercheurs en SHS ; quand un article était un peu douteux, on le leur envoyait pour avoir leur avis avant de le mettre en ligne.

David Rochefort : Et au niveau institutionnel ?

Daniel Charnay : Dans cette aventure des archives ouvertes lancée par le CNRS, deux EPST nous ont rejoints assez vite : l'Inserm et l'Inria. L'Inria a été et est toujours le bon élève de l'archive ouverte. C'est le seul qui, est arrivé à 100 % de sa production, en texte intégral, dans l'archive. Ils ont une vraie politique, grâce à des directeurs généraux qui ont adhéré au projet, et surtout à Jean-Pierre Verjus, qui était directeur de l'information scientifique et de la communication.

L'Inserm a dit « oui, pourquoi pas, mais à condition qu'on soit connecté à PubMed Central » (qui est l'équivalent d'arXiv pour les sciences de la vie). On les a donc contactés pour leur proposer d'être un frontal vers eux, comme on était un frontal vers arXiv. On a fait la connexion avec PubMed. Mais on n'a jamais atteint dans aucune discipline et dans aucun EPST, y compris le CNRS, l'exhaustivité du texte intégral.

J'ai passé deux ou trois années à courir toutes les universités, les EPST, à faire des présentations, à expliquer pourquoi il fallait déposer les preprint. La grosse réticence des chercheurs, c'était de dire « si je mets un preprint, la revue ne va plus accepter mon article ». Il est vrai que quelques revues faisaient cela – les revues de chimie par exemple. Si l'article était diffusé en preprint, c'était fini.

Il a fallu affronter toutes les peurs : ceux qui disaient « je n'ai pas le droit » ; ceux qui disaient « je ne serai pas

© CNRS Éditions | Téléchargé le 24/09/2023 sur www.cairn.info (IP: 90.66.145.31)

publié si je le fais » ; ceux qui étaient persuadés que le CNRS avait développé HAL dans un but unique d'évaluation. Certains sont très virulents sur ce thème, d'autant que, récemment, la ministre de l'Enseignement supérieur et de la Recherche a très largement soutenu HAL. Ils ressentent ça comme une perte de liberté : on aurait fabriqué le parfait outil pour « fliquer » les chercheurs. Certains sont dans cet esprit-là, malheureusement. S'ils connaissaient l'histoire de l'origine de HAL… Le CNRS était l'EPST qui s'en souciait le moins. Catherine Bréchignac a donné son accord à Franck Laloé pour qu'on se lance, mais sans plus : si ça marchait, tant mieux, sinon tant pis. On n'avait pas d'argent. J'ai démarré le labo avec trois ingénieurs à peine… Si l'IN2P3 n'avait pas été là pour me donner des machines, il n'y aurait pas eu HAL. Ça a changé, le ministère a récemment débloqué plus d'un million d'euros pour l'année. Le vent a tourné.

David Rochefort : Vous semblez regretter une dérive bibliométrique des archives ouvertes. Pouvez-vous nous en dire plus ?

Daniel Charnay : Je trouve cela catastrophique. Je me bats contre cela. La personne la plus convaincue qu'il fallait du texte intégral, c'était Franck Laloé. Moi, j'ai eu une très mauvaise idée, que je regrette. Pour essayer de faire venir les gens dans HAL et les fidéliser, j'ai inventé les portails. Si les gens se reconnaissaient à travers une page qui leur était dédiée, ça marcherait mieux. On a donc créé des portails d'université – sans même leur demander si elles étaient d'accord. On a mis leurs logos et affiché les publications de cette université. Ça a un peu marché.

Et puis Franck Laloé s'est dit : « si on avait l'exhaustivité de la production du laboratoire, ce serait formidable. En attendant que les gens mettent le texte intégral, on va donc leur permettre de mettre une notice. » Ce jour-là, on a trahi l'esprit de HAL. Je suis parti en 2011. Jusqu'à mon départ, on a essayé d'endiguer un peu le flot des notices, mais après… Il y a maintenant quatre fois plus de notices que d'articles en texte intégral. La situation est ubuesque puisqu'en parallèle, le Coso (comité pour la science ouverte) a été créé par le ministère (qui prend la suite de la Bibliothèque scientifique numérique). Comme on n'est jamais en mesure de savoir combien d'articles sont publiés en France, on a créé un outil pour recenser l'exhaustivité de la production scientifique française.

Cet outil, s'appelle Conditor, et vient de démarrer. Il s'auto-alimente dans HAL et il était, à ma connaissance, censé s'auto-alimenter auprès des grands éditeurs. En réalité, à ce que j'en sais, l'essentiel de Conditor aujourd'hui vient de HAL. Si ça marchait bien, on pourrait envisager une complémentarité : HAL leur donnerait les portails avec les notices et ne conserverait que articles avec texte intégral. On verra ce que nous réserve l'avenir.

Dans l'ensemble, je trouve que cette dérive est triste. On a le soutien du ministère, toutes les institutions ont signé un accord autour de Hal, il y a maintenant de l'argent, des postes, il y a tout ce qu'il faut pour faire quelque chose de bien, et on fait quelque chose qui s'éloigne de l'esprit original. Je trouve cela extrêmement dommage. Quand on voit dans la presse un article scientifique et qu'on cite une archive ouverte, on cite toujours le lien vers arXiv, jamais HAL.

David Rochefort : Quelle différence faites-vous entre le dépôt dans HAL et une publication dans une revue ?

Daniel Charnay : Avant on écrivait un article et il pouvait passer six mois ou un an avant qu'il ne soit publié. Là, on écrit un article, on dépose le preprint en archive ouverte, les gens de la discipline le commentent, font des remarques ; l'auteur corrige, modifie, fait plusieurs versions, et quand l'article est soumis à la revue, l'ensemble de la communauté l'a lu, et quasiment validé – et n'ira jamais lire l'article publié. En physique des particules, c'est flagrant. Les gens ont tous lu les articles de leur domaine bien

© CNRS Éditions | Téléchargé le 24/09/2023 sur www.cairn.info (IP: 90.66.145.31)

avant leur parution. La publication dans une revue sert seulement à donner une caution scientifique *a posteriori*.

David Rochefort : Pensez-vous qu'il y a un trop-plein d'informations accessibles ?

Daniel Charnay : Quelqu'un qui a un cancer va faire des recherches sur Internet, et va trouver des articles, par centaines. Il trouvera des articles traitant de solutions à sa maladie. On a vu ça : des gens vont voir leur médecin et disent « pourquoi je ne suis pas traité, j'ai lu une étude, etc. ». Une certaine partie de la population (des gens plutôt éduqués tout de même) va lire des articles scientifiques. Je ne sais pas si c'est bien ou pas.

En sciences humaines, je trouve des articles très intéressants. Après, c'est le problème de la connaissance, de la diffusion sur Internet de cette grande masse. Ce n'est pas propre aux archives ouvertes. Il faut éduquer les gens très tôt à trier l'information sur Internet et à apprendre à s'en servir. C'est le principal problème.

En revanche, pour les chercheurs, la question de cette masse d'information ne joue absolument pas. Les physiciens que je connaissais étaient abonnés au domaine scientifique qui les intéressait sur arXiv et ne lisaient essentiellement que ça. L'hypercroissance de l'information accessible n'a pas changé leurs habitudes de travail.

David Rochefort : Qu'entendiez-vous par l'expression « communication scientifique directe » ?

Daniel Charnay : Franck Laloé disait toujours qu'auparavant, les savants avaient des correspondances privées ; quand ils avaient une idée, ils l'envoyaient à quelques autres savants par la poste, la communication était directe. Franck voulait retrouver ce mode de communication. Mais là encore, celui qui a eu la grande idée, c'est Paul Ginsparg, quand il a créé arXiv. Quand il l'a fait, c'était avant le Web. Ça passait soit par courrier électronique, soit par FTP. Si vous regardez le site d'arXiv qui a peu évolué, c'est du Web des origines. Pas d'image, pas de bouton, c'est très minimaliste. HAL a une interface qui se veut un peu plus moderne.

Quand vous publiez chez arXiv, vous mettez votre article, votre nom, votre institution et c'est fini. Sur HAL, l'idée avait été de recueillir beaucoup plus de métadonnées. On demandait que tous les auteurs soient identifiés, et que leur appartenance soit renseignée. Quand vous avez un article avec 25 auteurs, c'était assez fastidieux, et cela représentait un frein. L'Inria avait donc inventé un petit système intelligent qui traitait l'article et arrivait à retrouver les auteurs, leur appartenance, etc.

David Rochefort : Quel avenir voyez-vous pour les archives ouvertes ?

Daniel Charnay : Je ne connais pas l'avenir des archives ouvertes. Certains éditeurs ont mis en place un nouveau système, où le laboratoire du chercheur paie pour qu'il soit publié et en libre accès. Si ce modèle marche, des archives ouvertes seront-elles encore nécessaires ? Oui, parce que sur le long terme, on peut envisager que certaines revues disparaissent sans rien conserver. Rien que pour l'archivage à long terme, elles ont donc un rôle à jouer… Les articles déposés dans HAL sont tous expédiés au Centre informatique national de l'enseignement supérieur (Cines), qui a un service d'archivage à long terme. On vérifie en permanence la lisibilité des articles. Si le pdf devient trop vieux et n'est plus lisible par les lecteurs actuels, l'article est automatiquement retraduit par le Cines. Ça, c'est important. L'idée de HAL, c'était d'avoir quelque chose de pérenne. C'est pourquoi HAL et le Cines sont partenaires depuis le début.

© CNRS Éditions | Téléchargé le 24/09/2023 sur www.cairn.info (IP: 90.66.145.31)