



HAL
open science

Des nains sur les épaules de géants : ouvrir la science en France

Marin Dacos

► **To cite this version:**

Marin Dacos. Des nains sur les épaules de géants : ouvrir la science en France. *Revue Politique et Parlementaire*, 2019. hal-02366604

HAL Id: hal-02366604

<https://hal.science/hal-02366604>

Submitted on 16 Nov 2019

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.

DES NAINS SUR LES ÉPAULES DE GÉANTS : OUVRIER LA SCIENCE EN FRANCE

La science ouverte est la diffusion des résultats de la recherche sans entrave et sans paiement. Elle représente une importante opportunité pour améliorer la qualité et le rayonnement de la recherche, mais aussi pour renforcer le lien entre science et société. Il s'agit d'abandonner les barrières à l'accès que constituent les abonnements aux revues scientifiques et de travailler à la diffusion de la matière première de la connaissance que sont les données de la recherche. Ce processus est mondial et concerne toutes les disciplines. La France a défini sa stratégie en 2018 en publiant le Plan national pour la science ouverte. Le public ciblé est bien sûr le monde académique à l'échelle de la planète, mais également la société au sens large : petites, moyennes et grandes entreprises, professions libérales, acteurs étatiques et collectivités territoriales, organisations non gouvernementales, acteurs de la médiation scientifique (comme Wikipédia) et citoyens.

Marin DACOS

Conseiller scientifique pour la science ouverte du Directeur général de la recherche et de l'innovation au ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation¹

Lorsque le *Journal des sçavans* naît à Paris en 1665, le monde de la recherche innove dans le domaine de la communication scientifique en inventant la première revue au monde. Jusque-là, les savants s'échangeaient des

lettres entre eux. Trois mois plus tard, Londres l'imite avec *Philosophical Transactions*, puis c'est l'Italie avec le *Giornale*

1 - Marin Dacos a fondé la plateforme OpenEdition en 1999 et est lauréat de la médaille de l'innovation du CNRS (2016).

de' letterati, puis l'Allemagne avec *Acta eruditorum Lipsiensium*. Aujourd'hui, on considère qu'il y a entre 50 000 et 100 000 périodiques scientifiques dans le monde. Le succès de cette innovation a contribué fortement à la construction de la science contemporaine. On lui doit, en tant que medium, une accélération et une amplification de l'impact des écrits des chercheurs.

En 1989, le Web est inventé par Tim Berners Lee au CERN, haut lieu de la recherche scientifique mondiale. Il constitue là aussi une innovation scientifique radicale, destinée à simplifier et accélérer le partage d'informations à l'échelle de la planète. Comme pour l'invention de la revue au XVII^e siècle, l'idée ne suffit pas à la propagation et à la maturation de l'innovation : il a fallu plusieurs siècles et des itérations pour inventer le système de l'édition scientifique qui est, de nos jours, arrivé à maturité. Il a fallu pour cela un processus de formalisation et de structuration, avec des métiers, des acteurs et des formes : l'évaluation par les pairs en double aveugle, la note de bas de page, la bibliographie, le résumé, etc., toutes formes qui n'existaient pas initialement ou sous des formes rudimentaires.

L'apparition du Web permet une évolution aussi radicale en s'appuyant sur une idée simple et géniale, permettant de diffuser les idées à la vitesse des élec-

trons ou des photons. Elle constitue le premier levier de la science ouverte. Le deuxième levier consiste à abandonner le système des abonnements, qui était en phase avec le système éditorial du papier, mais qui bride considérablement le système éditorial numérique. La science ouverte est ce concept qui permet de diffuser à tous, sans paiement, les publications scientifiques, mais aussi d'ouvrir les données de la recherche qui sous-tendent ces publications. Pour les données, l'usage fréquent d'un stockage local et fragile, et l'absence de tradition de partage ; pour les publications, la persistance de revues sous abonnement ; constituent des usages hérités peu adaptés au contexte numérique. La science ouverte a pour projet de dépasser cette situation.

La science ouverte présente de nombreux avantages. L'intérêt est d'abord politique : la science, essentiellement financée par la sphère publique, doit revenir au public, à tous les publics. L'intérêt est ensuite scientifique, car en faisant tomber les barrières à l'accès, l'accès universel aux publications améliore la qualité de la science. Elle présente, enfin, un avantage de société : une science plus ouverte est aussi plus accessible à la société. À l'heure où le discours scientifique ne peut plus s'imposer à la société par le seul pouvoir de l'argument d'autorité, ce processus d'ouverture est une opportunité pour replacer la recherche au centre

de la cité. Face aux enjeux majeurs que constituent les objectifs de développement durable de l'ONU, de biodiversité, de changement climatique, des transformations sans doute rapides et fortes sont nécessaires à l'échelle de la planète. La science ouverte constitue donc une opportunité exceptionnelle pour faire face aux défis du XXI^e siècle. Pourtant, le chemin est encore long, car la transition vers le nouveau paradigme ne fait que commencer et c'est à ce défi que se sont attaqués la loi pour une République numérique d'octobre 2016 ainsi que le Plan national pour la science ouverte annoncé par la ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation en juillet 2018.

OUVRIRE LES PUBLICATIONS SCIENTIFIQUES

UNE CRISE DE L'ACCÈS AUX PUBLICATIONS

De facto, les publications ne sont guère plus accessibles qu'au temps du support papier et de son nécessaire routage et acheminement en autant de lieux de lecture physique. Étrange paradoxe, alors que les publications n'ont jamais été aussi nombreuses, nous vivons une crise sans précédent de l'accès aux publications

scientifiques. Peter Suber, le spécialiste de la science ouverte à Harvard, l'a mentionné dans son ouvrage devenu classique, *Qu'est-ce que l'accès ouvert ?*

« En 2008, la bibliothèque de Harvard était abonnée à 98 900 périodiques, contre 73 900 pour l'université Yale. La même année, la bibliothèque de recherche indienne la mieux financée, celle de l'Indian Institute of Science, était abonnée à 10 600 périodiques. Plusieurs bibliothèques universitaires de l'Afrique subsaharienne ne sont abonnées à aucune revue scientifique payante. Leurs lecteurs n'ont donc pas accès à ce type de revue, à l'exception des abonnements offerts par quelques éditeurs conventionnels. »

Trop de chercheurs ont encore tendance à penser que tous leurs confrères ont accès à toute la production scientifique dès lors qu'ils sont assis à leur poste de travail dans leur bureau à l'université. Mais ce n'est pas le cas : dans le monde au sein duquel nous vivons, les inégalités d'accès sont abyssales, même en Occident, même dans des grandes universités, même dans les métropoles. Comprendre cela, c'est mesurer l'ampleur de la fracture de l'accès aux productions scientifiques dans l'univers académique, et par conséquent en mesurer l'ampleur – démultipliée – hors des campus, c'est-à-dire dans les petites et moyennes entreprises, dans les organisations non

gouvernementales, dans les collectivités, dans les écoles et jusque dans les foyers... Cet inventaire à la Prévert revient à construire une cartographie de l'ignorance scientifique...

Le magazine scientifique *Science* a confirmé ce diagnostic en 2016 en étudiant l'usage du site pirate Sci-Hub. Dans un article au titre provocateur, mais s'appuyant sur des données massives inédites, *Science* pose une question simple : « Qui télécharge des articles piratés ? ». La réponse, en lettres majuscules, est : « TOUT LE MONDE. Dans les pays riches comme les pays pauvres, les chercheurs se tournent vers le site Sci-Hub ».

Une étude qualitative reposant sur des entretiens avec de jeunes chercheurs dans plusieurs pays (Royaume-Uni, États-Unis d'Amérique, Malaisie, Chine, France) montre que ce type d'usage est extrêmement répandu (un quart des jeunes chercheurs l'utilisent régulièrement), avec une intensité semble-t-il plus forte encore en France qu'ailleurs. L'étude montre aussi que l'usage est lié à la fois à des problèmes d'accès (absence d'abonnements) et à des questions pratiques (complexité des systèmes d'authentification pour les campus abonnés).

Une étude portant sur six pays d'Amérique latine montre que 20 % des étudiants de ces pays utilisaient Sci-Hub en 2017. La Colombie et la Bolivie at-

teignaient même des niveaux supérieurs à 34 %, tandis que l'Argentine et le Paraguay avaient des taux très faibles (moins de 6 %).

En 2019, *The Lancet* a mené une étude des téléchargements de l'année 2016 portant sur la littérature médicale. Elle montre que les pays qui consultent le plus Sci-Hub étaient alors, en valeur absolue : l'Inde, la Chine, les États-Unis et l'Iran. Mais en valeur relative de la production scientifique nationale, ce sont les pays du milieu de classement (*lower-middle-income* et *upper-middle-income*) qui développent l'usage pirate le plus intensif : le nombre de téléchargements par rapport au nombre de publications du pays est le plus élevé dans les pays intermédiaires. Les pays les plus riches ont moins besoin de Sci-Hub, même s'ils en ont un usage intense, et les pays les plus pauvres (*low income countries*) semblent dans l'incapacité de profiter de cette manne de connaissances, sans doute en raison d'infrastructures scientifiques, médicales et techniques trop peu développées. L'enseignement principal de cette étude est que, en plus des grands pays scientifiques, la Russie, l'Algérie, le Brésil, la Turquie, le Mexique et l'Inde mobilisent considérablement Sci-Hub et s'en servent pour dépasser la crise de l'accès aux publications médicales.

Le cas de Sci-Hub est exemplaire, non parce qu'il constituerait un modèle à suivre,

Usage de Sci-Hub dans le domaine de la santé. Nombre de téléchargements illégaux par publication scientifique produite par chaque pays



Source : *The Lancet*.

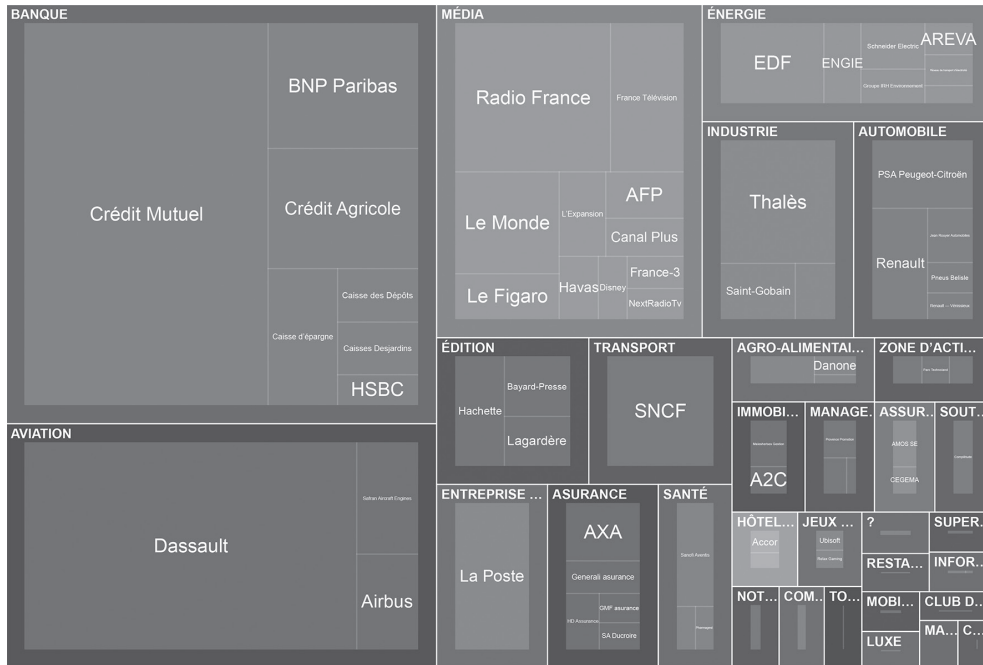
mais parce qu'il permet de quantifier et de cartographier la crise de l'accès aux résultats de la recherche scientifique. Sci-Hub est le symptôme de la maladie de l'accès qu'essaie de guérir l'accès ouvert légal.

Cette crise de l'accès est d'autant plus préjudiciable que nous avons désormais toutes les connaissances pour démontrer les bénéfices de l'ouverture des publications.

En effet, l'accès ouvert augmente les usages de consultation (Bacache-Beauvallet, 2015) et les citations. Des études en cours montrent également une diversification des usages, avec appropriation par des usagers éloignés de la communauté académique.

Les recherches en cours autour des usages d'OpenEdition montrent que cette plateforme française de sciences humaines et sociales est consultée par des entreprises du secteur de l'aéronautique (Dassault), de la banque (Crédit Mutuel), de l'assurance, de l'automobile, de l'énergie (Areva, Schneider Electric), et, ce n'est pas mineur pour la qualité du débat public, par les médias (Agence France Presse, Groupe France Télévision). Dans le secteur public, au-delà des campus universitaires, on identifie aussi des usages dans les administrations centrales et territoriales (Direction interministérielle du numérique et du système d'information et de communication de l'État, ministère de l'Aménagement du Territoire, de l'Équipement, Ville de Paris...).

Accès identifiés par des entreprises dans les 5 000 IP consultant le plus OpenEdition



Grâce à l'accès ouvert, la plateforme OpenEdition, spécialisée en sciences humaines et sociales, touche des entreprises dans de nombreux secteurs d'activités. Au Crédit Mutuel, on identifie au moins 99 000 accès sur la période 2017-mi 2019, portant principalement sur 18 000 articles différents. C'est similaire à la consultation de l'université de Lausanne. Publication à venir².

UN MOUVEMENT QUI PUISE SES RACINES DANS LES VALEURS SCIENTIFIQUES ET LES BESOINS DE LA RECHERCHE

La question de l'accès ouvert aux publications est ancienne : c'est en 1991 que les physiciens se sont dotés d'ArXiv, la première archive ouverte. Ils inauguraient ainsi la première décennie de la science ouverte (1991-2003), qui correspond à

un foisonnement d'initiatives pratiques, portées par des chercheurs, qui s'efforcent d'accélérer, simplifier et améliorer la communication scientifique entre pairs. Cette décennie est marquée par un fort prag-

2 - Ce travail a bénéficié d'une aide du gouvernement français au titre du Programme Investissements d'Avenir, Initiative d'Excellence d'Aix-Marseille Université - A*MIDEX. Aix-Marseille Univ, Avignon Université, CNRS, EHESS, CLEO, Marseille, France

matisme. Les physiciens sont frustrés par la lenteur du système éditorial, qui met des mois, parfois des années, entre le moment où l'article est soumis à une revue et le moment où celui-ci est publié. Ces derniers ont donc commencé à s'organiser en s'envoyant entre eux, par courriel, les PDF de leurs articles avant leur publication, en général au moment de leur soumission. Ce système ressemble étrangement à celui utilisé par les chercheurs au XVII^e siècle, avant l'invention de la revue... Comme au XVII^e siècle, les chercheurs ont constaté que ce système d'échange individuel était peu efficace. Installer un serveur central à Los Alamos, sur lequel pourront se connecter à tout moment tous les collègues qui le voudront, serait bien plus confortable. C'est l'origine de la création d'ArXiv : résoudre un problème pratique de lenteur et de partage. À la fin de la décennie, ce sont plutôt des plateformes de revues en accès ouvert qui sont créées par des chercheurs. Au Brésil : Scielo (1998), au Canada : Erudit (1998), en France : OpenEdition (1999, sous le nom Revues.org). Les économistes se sont, de leur côté, dotés de leur propre archive ouverte à l'échelle internationale (REPEC, 1997), tandis que le CNRS créait l'archive ouverte nationale de la France (HAL, 2001). Sont également créés des logiciels libres pour produire des archives ouvertes (Eprints, 2000 ; Dspace, 2002) ou pour héberger des revues (Open Journals Systems, 1999 ; Lodel, 2001).

Mais ces changements ne font pas basculer le système entièrement et la deuxième décennie est plus politique. Elle démarre avec la pétition initiée en 2000 par Harold Varmus (prix Nobel de médecine, ancien directeur du NIH), Patrick O. Brown et Michael Eisen, demandant aux éditeurs de publier leurs revues en accès ouvert, ou au moins en accès ouvert six mois après la publication. À l'issue de cette pétition, ils ont lancé PLOS, *Public Library of Science*, créant plusieurs revues en accès ouvert en biologie, toutes devenues importantes et reconnues. Entretemps, la Déclaration de Budapest (2002) avait conçu un programme politique, forgé le terme « accès ouvert » et constitué l'embryon d'une communauté militante au service de l'accès ouvert. Les initiatives, petites et grandes, couronnées de succès ou avortées, se multiplient. Elles sont portées par cette conviction que les abonnements sont un modèle anachronique, qui constitue un obstacle au progrès de la recherche et de la société.

Vient la décennie qui permet à la science ouverte de s'inscrire à l'agenda des politiques publiques. Celle-ci démarre par la recommandation européenne de 2012. L'Union européenne, qui dépense 80 milliards d'euros en recherche, est déçue par l'impact de ces importants efforts continuentaux. En effet, les accès des organisations de taille petite et moyenne, qu'elles soient privées ou publiques, sont difficiles, ne serait-ce qu'en raison du coût de transaction, le prix étant un obstacle supplé-

mentaire. Les PME, les ONG, les collectivités et les particuliers sont quasiment privés de cette masse de connaissances financées par l'Union. Pour résoudre ce problème, l'Union européenne rend obligatoire la diffusion en accès ouvert de tous les résultats qu'elle finance. La recommandation invite les États membres à faire de même pour les recherches qu'ils financent... sans grand succès. Constatant que le sujet n'avancait pas, l'*Amsterdam call for action on open science* de 2016 invitait chaque pays membre à se doter d'un plan national pour la science ouverte, ce qu'ont fait les Pays-Bas (2016) puis la France (2018), l'Irlande et la Suisse (2019). La France s'était dotée dès 2016 d'une législation accordant un nouveau droit aux chercheurs, celui de diffuser en accès ouvert leurs articles, six ou douze mois après la publication initiale, même après en avoir cédé les droits exclusifs à un éditeur. La Belgique s'est dotée en 2018 d'une législation similaire, avec cette fois une obligation de dépôt dans une archive ouverte d'établissement, avec les mêmes durées d'embargo que la loi française de 2016. Le Plan national français rend par ailleurs obligatoire la diffusion en accès ouvert de toute production scientifique financée par projets. Ces mesures parachèvent un mouvement entamé en 2008 avec l'obligation imposée par les puissants National Institutes of Health (NIH, Instituts américains de la santé), qui financent la recherche médicale et biomédicale aux États-Unis d'Amérique

à hauteur de plusieurs dizaines de milliards de dollars par an.

À l'échelle mondiale, les résultats de toutes ces politiques publiques sont déjà importants. En septembre 2019, le répertoire listant l'ensemble des revues en accès ouvert dans le monde, le DOAJ (Directory of open access journals) dénombrait 13 744 revues entièrement en accès ouvert, publiées dans 130 pays. La Confédération des archives ouvertes dans le monde (COAR) regroupait plus de 140 organisations sur les cinq continents. Tous les jours, l'accès ouvert gagne du terrain. Mais l'ouverture des publications, aujourd'hui bien entamée, mais inachevée, ne suffira pas.

LES DONNÉES DE LA RECHERCHE, MATIÈRE PREMIÈRE DE LA CONNAISSANCE SCIENTIFIQUE

Lorsque le programme politique de la science ouverte rencontre la question des données, le problème change de dimension. Par comparaison, la question des publications peut paraître simple. Les données de la recherche n'ont pas, comme les publications, un statut toujours considéré comme public : elles sont souvent appréhendées comme des matériaux relevant de l'atelier du cher-

cheur. Ce statut particulier est lié à la dimension exploratoire de l'objet ainsi qu'à sa très grande variété et complexité technique. Pour toutes ces raisons, les données de la recherche n'ont pas atteint la même maturité formelle et organisationnelle que l'édition. Dès lors, organiser, structurer et ouvrir des données de recherche se révèle beaucoup plus compliqué.

LE POTENTIEL DES DONNÉES DE LA RECHERCHE

Il n'existe globalement que deux formes de publications scientifiques : les articles publiés dans les revues et les ouvrages. Dans le domaine des données de la recherche, qui sont la matière première de la connaissance, on dénombre bien plus que deux formes : il y a un écart considérable entre « un cliché de coléoptère prélevé à Madagascar, un spectrohéliogramme produit à Meudon, des informations sur les gènes d'une moisissure, les relevés météorologiques d'un vaisseau ayant traversé l'Atlantique au XVIII^e siècle ou l'enregistrement d'un dialecte rare » (Rémi Gaillard, 2014), sans compter les données du télescope spatial Hubble, du synchrotron Soleil, des données marines collectées par la flotte d'Ifremer ou les produits de simulations climatiques *via* le calcul à haute performance (HPC)... Cette diversité est à la fois technique, économique et juridique.

Certaines disciplines ont une longue tradition de publication de leurs données. Depuis longtemps, les archéologues publient les relevés de leurs fouilles, les juristes publient la jurisprudence en droit, les généticiens publient systématiquement leurs données génétiques, et les astronomes publient leurs observations. Prenons deux exemples, parmi d'autres, présentant l'utilité des données et l'intérêt d'en disposer même une fois une recherche terminée. Le chercheur Joseph Charles Farman avait pris l'habitude d'enregistrer la concentration d'ozone en Antarctique. En 1980, il est surpris de noter que les mesures apportées par son spectrophotomètre indiquent une baisse de la concentration d'ozone. En comparant ses résultats avec ceux de la Nasa, il conclut que ses mesures sont défectueuses et ne s'en inquiète pas. Mais lors d'une nouvelle expédition en 1984, avec de nouveaux instruments, il obtient des mesures similaires... Interrogée à nouveau, la Nasa lui fournit des données qui laissent penser que la concentration d'ozone n'a toujours pas baissé... Finalement, ses instruments disaient vrai depuis le début, le fautif étant un algorithme de lissage des résultats mis en œuvre par la Nasa. En remontant aux données non traitées par l'algorithme, le trou était bien présent depuis des années dans les mesures de la Nasa, mais devenu invisible. C'est ce retour aux données originelles et la confrontation scientifique qui ont permis de mettre en évidence un danger réel, auquel a répondu le protocole

de Montréal en 1987. En sismologie, le principe du partage et de l'ouverture des données entre tous les chercheurs est acquis au point que toutes les publications scientifiques portant sur la structure de la terre globale, sur la terre profonde ou sur la sismicité, sont basées sur des données ouvertes, qui sont partagées sous forme numérique depuis le milieu des années 1980. Ainsi, une multiplicité de points d'observations permet d'avoir une appréhension globale des phénomènes sismologiques.

Cependant, ces exemples ne concernent pas toutes les disciplines.

UNE INÉGALE GESTION DES DONNÉES DE LA RECHERCHE

De nombreuses disciplines n'ont pas de gestion collective des données de la recherche. Les données sont gérées à des échelles variables, de l'ordinateur du chercheur au serveur local de son centre de recherche. Par ailleurs, il y a beaucoup de disciplines qui ne produisent publiquement que leurs conclusions, et pas leurs matériaux. Dans ce contexte, la réutilisation des données et la reproductibilité des résultats scientifiques sont souvent limitées. Plusieurs analyses, de 2011 et 2012, dans le domaine de la médecine estiment que seulement 11-25 % des études sont reproductibles, faute de données... en d'autres mots, à cause de la perte des don-

nées ou parce que les données ne sont pas ou plus accessibles. Une analyse de publications en morphologie révèle une baisse progressive de la disponibilité des données sur une période de vingt ans (1991-2011). Deux ans après la publication, la plupart des données sont encore disponibles aux autres chercheurs. Mais vingt ans plus tard, 80 % des données ont été perdues, avec une baisse moyenne de 17 % par an...

D'une façon plus générale, la disponibilité des données n'est pas du tout suffisante. Il faut, en effet, qu'elles obéissent aux principes FAIR pour qu'on puisse les trouver, les comprendre et les réutiliser. Les données devraient donc à terme être (F) faciles à trouver, (A) accessibles, (I) interopérables et (R) réutilisables. L'Union européenne a commandé une étude pour estimer le coût d'avoir des données non conformes à ces principes FAIR. Les résultats, publiés en 2019, montrent que l'absence de respect de ces bonnes pratiques représenterait un coût de 10 milliards d'euros par an... soit 3 % des dépenses de recherche européennes, ou 4 fois le budget de l'ERC - European research council. L'étude se base sur un certain nombre d'indicateurs permettant de cerner les coûts et bénéfices liés à la structuration et à l'ouverture des données de recherche. Par exemple, les auteurs postulent qu'une partie importante des rétractations, c'est-à-dire des articles publiés dans des revues puis retirés quelques années plus tard en raison d'erreurs ou de fraudes, s'explique par une mauvaise ges-

tion des données sous-jacentes à l'article, rendant difficile leur évaluation et leur vérification. Ils mesurent également le temps passé à reconstruire des données déjà produites par ailleurs, mais inaccessibles, perdues ou inutilisables. Ils évaluent aussi les freins à l'accès aux données en termes de coûts de transaction, d'obstacles à l'interdisciplinarité, de freins à l'exploitation économique. Même si une telle étude peut être méthodologiquement discutée, elle permet de mesurer que produire des données coûte de l'argent et que mal les gérer en coûte également, sans doute beaucoup plus. Rappelons, par ailleurs, que certaines données nourrissent un certain nombre d'activités économiques. À l'échelle européenne, on considère que le chiffre d'affaires des industries s'appuyant sur des données s'élève à 300 milliards en 2016 (en croissance de 5,03 % par rapport à 2015), avec une projection à 4 % du PIB européen en 2020 (739 Mds €)³. Les données de la recherche en font partie.

Si la question est économique, elle est également juridique, organisationnelle et culturelle. D'un point de vue juridique, la loi pour une République numérique d'octobre 2016 a mis en place le principe de l'ouverture par défaut des données administratives, c'est-à-dire de données produites dans le cadre de missions de service public. Les données produites par les cher-

cheurs étant juridiquement des données administratives, celles-ci sont concernées par l'ouverture par défaut. Si la loi n'entérine pas explicitement le principe « aussi ouvert que possible, aussi fermé que nécessaire », elle est pourtant fidèle à ce principe, puisque de nombreuses exceptions au principe d'ouverture sont prévues⁴. Mais, finalement, ce qui manque le plus, c'est de la maturité organisationnelle et culturelle des communautés scientifiques.

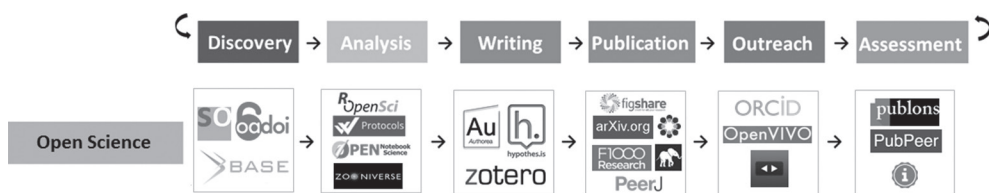
LE CYCLE DE VIE NUMÉRIQUE DE LA RECHERCHE

En dehors de certaines communautés très structurées autour des données de la recherche, il serait en effet illusoire de croire que les chercheurs disposent des services numériques adaptés à leur besoin, dans un contexte ergonomique favorable et dans un environnement sécurisé techniquement et juridiquement. Actuellement, dans beaucoup de disciplines, les chercheurs utilisent des dizaines d'outils dispersés tout au long du cycle de vie de la recherche. La fragilité de cet écosystème et son rattachement

4 - Le principe d'ouverture n'est pas applicable dans le cas de données personnelles sans consentement, de données présentant des risques pour la sécurité publique ou la sécurité de l'établissement, le secret de la défense nationale, le secret professionnel, le secret des affaires, les secrets industriels et commerciaux. Les données présentant des risques pour la protection du potentiel scientifique et technique de la nation.

3 - <https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy>

Chaîne de services et logiciels relevant de la science ouverte, selon Bianca Kramer et Jeroen Bosman



Chaîne de services et logiciels proposés par Elsevier, n°1 mondial de l'édition scientifique



AGIR POUR LA SCIENCE OUVERTE

LE PLAN NATIONAL POUR LA SCIENCE OUVERTE

Face à ces enjeux, la France a décidé d'engager une ambitieuse réforme en faveur de la science ouverte. Annoncé en juillet 2018 par la ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation, Frédérique Vidal, le Plan national pour la science ouverte comporte trois volets. Le premier concerne l'ouverture des publications ; le deuxième volet concerne la structuration et, quand c'est possible, l'ouverture des données de la recherche ; le troisième a pour vocation de produire une transformation structurelle, à l'échelle locale (en passant par les écoles doctorales), à l'échelle

nationale (en sollicitant les établissements d'enseignement supérieur, de recherche et d'innovation pour qu'ils se dotent de plans pour la science ouverte) et à l'échelle internationale (en s'appuyant notamment sur la structuration européenne que constitue le projet European Open Science Cloud, EOSC).

Il est bien entendu un peu tôt pour mesurer l'impact de ces mesures, dont la plus emblématique est la mise en place d'une obligation nationale d'ouverture des publications pour toutes les recherches qui sont financées par projets avec de l'argent public. Cependant, le plan prévoyait la mise en place d'un baromètre de la science ouverte, dont les premiers résultats montrent que mi-2019, les publications scientifiques françaises de l'année 2017 étaient en accès ouvert à 41 %. Les résultats montrent

une extrême variation disciplinaire, dans les taux d'accès ouvert : entre 29,7 % pour les sciences sociales et 61,6 % pour les mathématiques, avec de faibles scores pour la chimie (31,3 %) et de bons scores pour la physique et l'astronomie (55,5 %). Plus intéressant encore : alors que la médecine et les sciences sociales sont diffusées en accès ouvert grâce à des revues en accès ouvert, les mathématiques et l'informatique sont en accès ouvert essentiellement grâce à des archives ouvertes dans lesquelles les auteurs déposent individuellement (sur le modèle d'ArXiv et de HAL). Ces résultats confortent le plan national dans sa volonté de jouer sur tous les leviers existants pour

ouvrir la science, sans privilégier une voie particulière : il est préférable de s'adosser sur les dynamiques initiées par les disciplines et leur culture, bien souvent internationale, afin de s'appuyer sur ces forces plutôt que de chercher à les contrecarrer.

Le plan national a décidé de créer le Comité pour la science ouverte, chargé de coordonner l'action nationale en s'appuyant sur les communautés scientifiques. Plutôt que d'imposer aux communautés des règles qui seraient hors sol, orthogonales à leurs cultures et usages existants, le plan fait le pari d'aider à l'organisation des communautés scientifiques. C'est l'esprit

Résultats provisoires du Baromètre de la science ouverte développé par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (septembre 2019). À paraître.

Baromètre français de la Science Ouverte

Mesurer de façon régulière l'évolution de l'accès ouvert en France à partir de données fiables, ouvertes et maîtrisées



de l'appel ANR Flash Science Ouverte sur les données de la recherche, qui a été lancé en 2019 (dotation de 2,3 M€). Le Fonds national pour la science ouverte a par ailleurs été créé le 17 juillet 2019, comme l'avait annoncé le plan, et s'apprête à lancer son premier appel à projets. Pour l'année 2019, le fonds est doté de 3,1 M€ et il peut accueillir des contributions volontaires de la part d'institutions françaises ou étrangères, privées ou publiques, qui souhaitent soutenir la science ouverte. Au total, le Plan national pour la science ouverte est doté de 5,4 M€.

DES NAINS SUR LES ÉPAULES DE GÉANTS... AU SERVICE DE LA SOCIÉTÉ

Le paradigme de la science ouverte vise à renforcer l'idée que la recherche est l'affaire des nains juchés sur les épaules de géants, c'est-à-dire de chercheurs qui s'appuient sur le travail collectif de la communauté scientifique, incarné par ses publications, ses données et ses méthodes. Ce programme repose sur trois principes. Tout d'abord, rendre au public ce qui a été financé par le public. Ensuite, développer une recherche de meilleure qualité grâce à la réduction, autant que possible, des effets de voile imposés par les difficultés d'accès aux publications et aux données. Enfin, une recherche plus ouverte est plus accessible à la société, et donc susceptible de nourrir celle-ci de façon plus efficace. En-

gagée depuis trente ans, la transformation de la recherche vers plus d'ouverture est en train de se généraliser à l'échelle mondiale, notamment à travers des plans nationaux pour la science ouverte, mais aussi à travers des changements de culture et de pratiques dans les communautés scientifiques. L'ouverture progresse : à l'échelle de l'Union européenne, 35,7 % de la production de 2017 serait en accès ouvert (source : Open science monitor de l'Union européenne). Est-ce qu'on peut déjà mesurer les impacts sur la société, ce qui est la promesse ultime du paradigme de la science ouverte ? Cela est bien entendu difficile à évaluer. Mais les études portant sur la façon dont les articles de Wikipédia sont alimentés par le savoir académique démontrent, de façon assez nette, l'efficacité de l'ouverture pour l'enrichissement de la compréhension du monde par les citoyens et les organisations. La version anglophone de Wikipédia cite déjà 300 000 articles scientifiques et 18 000 ouvrages en sciences humaines et sociales (Kousha, 2017). Surtout, il est démontré qu'un article en accès ouvert a 47 % plus de chances d'être cité dans Wikipédia qu'un article en accès restreint (Teplitskiy, 2017). Ces études montrent que les nombreux contributeurs de Wikipédia utilisent la littérature scientifique, et que ceux-ci citent en priorité les contenus auxquels ils ont facilement et légalement accès grâce à l'accès ouvert. Existe-t-il démonstration plus éclatante de la capacité de l'accès ouvert à dépasser les frontières du monde académique pour toucher la société ? ■