



Contexte Rencontre avec **Nicolas Fressengeas**, vice-président en charge du numérique, des données et de la Science Ouverte à l'Université de Lorraine qui a rejoint le Conseil d'orientation de l'Ofis en juin 2022. Il nous explique les enjeux cruciaux en 2023 à ses yeux, à la croisée de l'intégrité scientifique et de la science ouverte.



Nicolas Fressengeas



Professeur de physique, vos fonctions actuelles sont toutes liées à l'ouverture de la science dans toutes ses dimensions et à la promotion des bonnes pratiques d'intégrité scientifique : quelle sont les motivations qui ont guidé vos choix ?

Nicolas Fressengeas : L'accès le plus libre possible aux résultats scientifiques et aux avancées de la connaissance en général a toujours été une préoccupation importante pour moi. Je me suis intéressé très tôt aux nouveaux modes de diffusion que les outils numériques ont permis de développer. J'ai commencé à déposer mes articles sur la plateforme d'archives ouvertes HAL dès 2006 et me suis ainsi inscrit naturellement dans le mouvement de la science ouverte. J'utilise aussi des logiciels libres, comme Linux depuis 15 ans. Pour ces raisons, l'Université de Lorraine m'a d'abord confié une première mission Science Ouverte en 2019 et j'ai été nommé vice-président en charge *du numérique, des données et de la science ouverte* en 2022. Quant à l'intégrité scientifique, les bonnes pratiques sont à la base de la démarche scientifique. Pourtant, en tant que directeur de laboratoire, j'ai été témoin de mauvais comportements. J'ai vu la montée des revues et conférences prédatrices de près. J'ai même failli me faire piéger, il y a 10 ans déjà. Aujourd'hui, je suis toujours stupéfait du sentiment d'impunité qui continue de régner chez certains chercheurs, certes rare mais qui fait des dégâts, et à quel point l'université elle-même peut parfois être démunie face à ce type de problèmes. Je pense que plus la science sera ouverte, moins les écarts à l'intégrité pourront être dissimulés.

Quels sont pour vous, aujourd'hui en 2023, les enjeux communs à l'intégrité scientifique et la science ouverte les plus importants ?

N.F.: Pour moi, les principaux enjeux correspondent aux quatre piliers du deuxième plan national pour la science ouverte. Le premier, la généralisation de l'accès ouvert des publications scientifiques (articles, livres, monographies), est bien avancé : le baromètre national montre que 62 % des publications de 2021 sont en accès ouvert, contre seulement 40% il y a quelques

En quatre traits

Aujourd'hui. vice-président en charge du numérique, des données et de la science ouverte à l'Université de Lorraine, chargé de mission Science Ouverte, volet international pour le Ministère de l'Enseignement Supérieur et de la Recherche, membre du Conseil d'orientation de l'Ofis.

Formation. Docteur en Physique de l'Université de Metz. Professeur de l'Université de Lorraine.

Expériences fondatrices : Directeur du laboratoire Matériaux Optiques, Photonique et systèmes (LMOPS) à l'université de Lorraine de 2017 à 2022. Il a été acteur de la transition pédagogique impliquant les techniques numériques à l'université de Lorraine. <https://orcid.org/0000-0002-5534-712X>

Ligne directrice : l'ouverture de la science, c'est-à-dire de l'ensemble des productions de la recherche (publications, données, codes et logiciels), au plus grand nombre et au service d'une recherche intègre.

années, et la progression est continue ! Il reste néanmoins que certains modèles de publication en accès libre font exploser les coûts pour les auteurs et que des dérives comme les revues prédatrices ou les usines à articles prolifèrent : nous avons donc à progresser collectivement sur les bonnes pratiques. Les deuxième et troisième piliers de la science ouverte à savoir *structurer, partager et ouvrir les données de la recherche* ainsi qu'*ouvrir et promouvoir les codes sources produits par la recherche*, sont directement liés par le décret du 3 décembre 2021 relatif aux questions d'intégrité, puisque ce dernier oblige les établissements à définir et mettre en œuvre une politique de conservation, de communication et de réutilisation des données et codes source, en lien avec la science ouverte. Enfin, le quatrième pilier qui consiste à *transformer les habitudes pour que ces trois pratiques deviennent les pratiques par défaut*, est dans la ligne de toutes les initiatives qui cherchent à créer des environnements de recherche plus favorables aux bonnes pratiques, comme la coalition européenne CoARA (*coalition for advancing research assessment*). CoARA vise à faire évoluer les pratiques de l'évaluation de la recherche notamment en diminuant la pression à la publication et en reconnaissant la diversité des productions scientifiques, comme les données de recherche. Un premier horizon est fixé à 5 ans mais ces changements de culture prendront probablement plus de temps.

La question des modèles qui font porter le coût de la publication en accès libre sur les auteurs (les APC, *article processing charges*) est particulièrement d'actualité : la première évaluation du coût de ces APC en France estime leur montant pour l'année 2020 à 30,1 millions d'euros. Un coût qui a triplé en moins de 10 ans. Qu'en pensez-vous ?

N.F.: Les projections sont encore plus préoccupantes : ces coûts pourraient doubler dans la prochaine décennie. Ce modèle, dit *gold*, n'est pas tenable ! Et sur le principe, il présente un sérieux problème. A partir du moment où l'auteur paye pour publier en accès libre, l'éditeur a tout intérêt à accepter son article ! On entre dans des relations commerciales avec des enjeux financiers énormes, qui peuvent générer toute sorte de dérives. Cela conduit par exemple à des pratiques questionnables : certains éditeurs proposent des réductions d'APC, parfois importantes, si l'on soumet son article avant une date donnée ou si l'on a rendu des services, éditoriaux ou d'expertise, à leurs revues. Ces articles, que l'on pourrait presque qualifier de soldés, sont-ils de moindre qualité ? On le voit, ces éditeurs peuvent avoir des démarches commerciales très agressives. A l'Université de Lorraine, nous avons mis fin à un abonnement global à l'ensemble des revues d'un gros éditeur, leurs commerciaux ont par la suite appelé tous les directeurs de laboratoires et de pôles scientifiques pour qu'ils fassent pression sur les services bibliothécaires. Résultat, les services des bibliothèques ont été accusés d'empêcher les chercheurs de travailler. Il n'est pas possible de cautionner un tel système : que le chercheur paie pour publier n'est pas une solution ! Nous encourageons donc nos enseignants-chercheurs à publier via d'autres modèles et venons de définir notre politique d'abonnement, en étant très vigilants sur certains accords globaux proposés par les éditeurs.

Tout modèle de publication de revue scientifique, même en ligne, exige des moyens et des compétences. Si les auteurs ne paient pas et les lecteurs non plus, qui paie ? Quel modèle préconisez-vous ?

N.F.: Pour moi, l'une des solutions c'est le modèle diamant : gratuit pour le lecteur, gratuit pour l'auteur et financé soit par des universités, soit par des États, voire par des mécènes. D'importants projets sont lancés au niveau européen (Action Plan for Diamond Open Access et OPERAS). Certes, un débat existe sur le sujet, et les modèles de financement restent à explorer. Mais, s'agissant d'une recherche scientifique publique dont quasiment toutes les étapes sont gérées par la communauté académique, il n'est pas illogique que la toute dernière étape, celle de la publication, revienne dans le giron académique. Historiquement, elle a en effet été externalisée, confiée à des acteurs privés,



parce que l'impression, l'édition, la distribution n'étaient pas considérées du ressort du monde académique. Aujourd'hui, avec les revues numériques, la question de la réappropriation par les acteurs académiques de cette étape-là me paraît complètement pertinente.

Vous l'avez rappelé, le décret de décembre 2021 lie la question de la conservation des résultats bruts de la recherche à l'intégrité scientifique et aux objectifs de la science ouverte. Cette nouvelle obligation suscite de nombreuses questions : quelles données publier, conserver ? comment ? où et qui pour le faire ?

N.F.: Ces interrogations sont légitimes. Néanmoins beaucoup de ces questions ont été largement défrichées et des réponses pratiques existent sur le portail de l'écosystème national au service du partage et de l'ouverture des données de recherche : [Recherche Data Gouv](#). Précisons d'abord qu'en termes d'intégrité, la publication et la conservation des données permet non seulement à toute la communauté de reproduire et de vérifier les résultats, mais permet aussi de détecter si des données ont été falsifiées, embellies... En la matière, la première et la plus simple des bonnes pratiques à mettre en œuvre est de déposer ses données dans un entrepôt de confiance dont [Recherche Data Gouv](#) proposera bientôt une liste, et qu'il référencera dans son catalogue. Il en découle que tout usage d'un jeu de données dans un travail de recherche doit être accompagné d'une citation de ce jeu de données, via son DOI propre. Dans certaines disciplines cette pratique est déjà bien ancrée, comme le révélera prochainement [le baromètre de la science ouverte](#). Bien sûr, toutes les disciplines ne le font pas de la même façon. Dans la mienne, par exemple, l'optique et la science des matériaux, ouvrir ses données ne serait-ce qu'au collègue du bureau voisin, ne va pas du tout de soi.

Publier ses données sur Recherche Data Gouv, c'est donc bien répondre à la nouvelle obligation ?

N.F.: Les données publiées sur [Recherche Data Gouv](#) sont conservées et les métadonnées sont publiques. C'est donc une solution pour répondre à l'obligation de conserver ses données de recherche. Pour être plus précis, on peut publier les données de deux façons: en mode ouverture et en mode partage. En mode ouverture, les données et les métadonnées associées qui décrivent ces données sont accessibles à tous, humains comme machines ou intelligence artificielle. En mode partage, seule les métadonnées sont en libre accès et les données sont accessibles sur demande.

Le cas des données directement associées à un article est le plus simple. D'autres cas posent plus question, faut-il publier des données qui n'ont pas encore fait l'objet d'un article associé ? A quel moment, selon quelles modalités ?

N.F.: Cela soulève deux points. Premièrement, tout comme on ne publie pas tous ses résultats de recherche, le choix des données à publier relève d'une politique scientifique. Et les plans de gestion des données doivent aider à développer cette politique. Deuxièmement, c'est toute la question de la reconnaissance du travail de production de données. Imaginons un exemple fictif un peu extrême: un chercheur développe un nouvel outil de cartographie mondiale de la température des océans, très performant, en temps réel à une échelle inédite. Faut-il qu'il publie ces données alors que, faute de temps, il n'a pas encore publié d'article? Dans l'intérêt général, la réponse paraît évidente. Mais pour que ce chercheur s'y retrouve et ne se sente pas dépossédé, ses données doivent exister comme une production scientifique en soi, prise en compte dans l'évaluation au même titre qu'un article, avec un DOI propre que l'on puisse citer en tant que tel.



Les logiciels, considérés comme résultats bruts de la recherche, doivent-ils être également publiés ?

N.F.: Pour les logiciels, on part de plus loin encore mais oui, l'idée est toujours la même: ouvrir l'ensemble des productions de la recherche, le fameux triptyque : publications, données et logiciels ou codes sources. Du point de vue de l'intégrité scientifique, publier les données, sans le logiciel qui sert à les traiter laisse une part du processus obscure. Pour les disciplines peu familières de l'univers du logiciel libre, les notions d'ouverture, de partage, de conservation de ces codes qui se déclinent de manière un peu plus complexe sont certes plus difficiles à appréhender. Mais grâce au projet d'archives mondiales des logiciels libres développé par INRIA depuis plus de 5 ans, *Software Heritage*, il existe déjà un entrepôt pour les conserver. En France, l'on peut ensuite déposer la version aboutie du logiciel dans HAL qui est connecté à *Software Heritage*. Là encore, les guides de la science ouverte expliquent très bien la marche à suivre.

Observez-vous des difficultés, des freins dans la mise en œuvre sur le terrain ?

N.F.: Le principal frein est le manque de temps. Une tâche comme la gestion de données vient s'ajouter à toutes les autres qu'un enseignant chercheur peut être amené à conduire (faire de la recherche, publier ses résultats, donner des cours, faire de la pédagogie nouvelle, gérer les filières, gérer des laboratoires, gérer un pôle scientifique etc.). Pour qu'elle soit prise en charge, il faut que le travail pour produire ces données, les mettre en forme, les rendre réutilisables, interopérables..., soit reconnu au même titre que les autres tâches et soit valorisé dans la carrière, d'une manière ou d'une autre. On en revient toujours à la question de la reconnaissance. Une autre difficulté tient aussi parfois à la réception de ces nouvelles lignes directrices. La demande d'un plan de gestion des données est parfois perçue comme une exigence administrative supplémentaire. Plus généralement, la normalisation de règles de bon exercice du métier est souvent mal accueillie dans nos milieux académiques. Proposer à un chercheur de changer un peu sa démarche n'est pas chose évidente. Mais l'université se doit, a minima, de lui offrir l'environnement qui lui permet de le faire (outils, formation, moyens) car je pense que nous avons tous à y gagner, en terme de fiabilité de nos résultats, de nos données, de notre démarche.

Un autre enjeu est la traçabilité de l'ensemble du processus pour identifier les problèmes et éviter leur propagation, par exemple éviter qu'un article rétracté continue d'être cité et encore moins utilisé : l'ouverture des données et les outils actuels permettent-ils de le faire ?

N.F.: Pas encore, mais c'est l'étape suivante et beaucoup de gens y travaillent. L'idée est d'ouvrir non seulement les articles, les données, les codes mais l'ensemble du processus de recherche. C'est par exemple le concept de *Research Nexus* porté par CrossRef*, organisme qui délivre les identifiants (DOI) pour les publications, qui propose d'attribuer à tous ces objets des DOI* de manière à les lier numériquement avec des métadonnées qui explicitent ces liens et la nature des documents. Ainsi, toute la vie d'un article de la première version à la version définitive - s'il a été rétracté, s'il a suscité des commentaires etc. - devient traçable et publique. Pour l'instant, concernant les rétractations, des métadonnées existent, et elles sont notifiées sur HAL, Zotéro. Mais on est loin d'une utilisation systématique. Le projet de Crossref œuvre dans ce sens ; il faudrait néanmoins veiller à éviter tout monopole et surtout ne pas oublier l'interconnexion avec les données et les codes.

Ces projets vont-ils dans le sens du mouvement de l'open reviewing, l'ouverture de la procédure de revue par les pairs, qui prend de l'ampleur ?

N.F.: Ils pourraient l'intégrer. Le fait que les rapports de revue par les pairs ne soient pas publics pose un certain nombre de problèmes. Aujourd'hui, si on me rapporte qu'un article a été publié dans une « grande » revue non pour sa qualité mais via des connaissances au comité éditorial, il m'est

impossible de le vérifier, le rapport de *review* étant confidentiel. Cette confidentialité a été pensée pour prévenir des pressions sur le *reviewer*, pour qu'un jeune chercheur puisse éventuellement critiquer le travail d'un chercheur plus confirmé par exemple. Mais l'inconvénient c'est l'opacité. Si le rapport devient consultable, lié par son propre DOI à l'article, et qu'il est de qualité, il n'y plus de problème. Une telle ouverture permettrait, si ce n'est de les éviter, au moins de rendre visibles certaines mauvaises pratiques, les rapports de relecture trop légers seraient faciles à repérer et l'absence de rapport, encore plus.

Enfin, la science ouverte vise à ouvrir au plus grand nombre les avancées de la recherche. Ne risque-t-elle pas d'accentuer la fracture numérique entre les pays les mieux dotés et ceux qui n'ont pas les moyens de publier en accès libre ?

N.F.: L'accès à la communication scientifique pour les pays aux plus faibles revenus est effectivement un enjeu crucial : l'ouverture des productions scientifiques -articles, données et logiciels- en garantit l'accès à tous et répond ainsi à une partie de l'enjeu. Mais qu'en est-il de l'autre volet : comment assurer la possibilité pour tous de publier les productions de la recherche en accès ouvert ? Certains modèles économiques de publication, comme les accords dits transformants, proposés par les grands éditeurs, ne sont pas de nature à favoriser l'équité, contrairement au modèle diamant évoqué plus haut. C'est toute la question, plus générale, des infrastructures de la Science Ouverte et de leur financement. L'Unesco peut avoir un rôle à jouer, comme elle le fait pour *Software Heritage*, ouvert au niveau mondial. Le risque est cependant réel, et en être conscient est le premier pas pour l'éviter.

Propos recueillis par Hélène Le Meur

Crossref est une organisation à but non lucratif de droit américain, enregistrée aux Etats-Unis sous le nom PILA (Publishers International Linking Association) qui compte 17 000 adhérents, éditeurs et autres institutions.

DOI pour digital object identifier (identifiant d'objet numérique).

