

Vive les données libres !

Permettre à tous d'accéder aux données brutes des chercheurs pour faire avancer le savoir : c'est l'idée défendue par Dominique Roche, qui milite pour une « science ouverte ».

Par Marine Corniou



C'

est pendant son doctorat en écologie, en 2013, que Dominique Roche a commencé à réfléchir à la notion de science ouverte. Il est tombé sur le sujet un peu par hasard, à la faveur d'une discussion organisée par le club de lecture du Département de biologie de l'Université nationale australienne, où il étudiait les poissons des récifs coralliens. Immédiatement, il s'est passionné pour la question : et si, au lieu de garder jalousement leurs données, les chercheurs les partageaient pour permettre à d'autres équipes de tirer parti de leur travail ? Le thème est délicat dans un milieu hautement compétitif.

Puis, en 2020, le biologiste obtient une bourse Marie-Sklodowska-Curie de la Commission européenne pour se consacrer à cette réflexion pendant trois ans, dans le cadre d'un postdoctorat réalisé conjointement à l'Université Carleton, à Ottawa, et à l'Université de Neuchâtel, en Suisse.

Le moment ne pouvait être meilleur : la pandémie a montré que la science ne va jamais aussi vite que lorsque ses artisans travaillent main dans la main. Même le G7, en juin dernier, a reconnu publiquement l'importance de renforcer la collaboration en matière de recherche et de promouvoir la science ouverte. Un signe que le vent tourne.

Québec Science : Qu'est-ce que la « science ouverte » ?

Dominique Roche : L'idée est simple : il s'agit de rendre plus accessibles le processus et les résultats de recherche.

Cela peut être fait de plusieurs façons. D'abord, en autorisant l'accès libre aux articles scientifiques. La recherche est financée par l'argent des contribuables, c'est donc illogique qu'elle soit ensuite publiée par des éditeurs privés derrière un mur payant ! Quand les gens réalisent cela, ils n'en reviennent pas.

La science ouverte encourage aussi le partage des jeux de données qui ont servi à la publication et celui du code utilisé par exemple pour les analyses statistiques.

QS Pourquoi est-ce important, selon vous ?

DR L'objectif est de favoriser l'avancement du savoir, plutôt que la compétition et la progression de carrière des individus. C'est d'autant plus important lorsqu'on fait face à des problèmes urgents et planétaires, comme les changements climatiques, la perte de biodiversité ou une pandémie.

C'est d'ailleurs pendant la crise de la COVID-19 qu'on a commencé à parler davantage du partage des données. Les génomes viraux ont ainsi été rendus publics et il y a eu plus de collaboration dans un domaine habituellement régi par



la course au brevet. Cela étant, c'était loin d'être parfait : une étude a montré que seulement 13 % des articles publiés sur la COVID-19 [entre décembre 2019 et avril 2020] incluaient des données de recherche. Cela m'a beaucoup surpris ; ce n'est pas encore une pratique acceptée de tous.

QS Quelle est la différence entre l'accès aux articles scientifiques et l'accès aux données ?

DR L'article scientifique, c'est une interprétation des données, qui repose en général sur des analyses statistiques spécifiques. C'est très superficiel. Il y a plein d'étapes entre la collecte des données et la publication, avec des possibilités d'erreurs ou encore des divergences d'opinions sur la façon dont on devrait analyser et interpréter les données.

Quand on a accès à l'article de recherche, on a des tableaux, des figures, mais c'est souvent incomplet et mal expliqué. Ce que les chercheurs veulent maintenant, c'est pouvoir évaluer le processus entre la collecte des données et l'interprétation, en ayant accès aux données primaires recueillies.

QS Le processus de révision par les pairs, qui consiste à faire revoir un article avant sa publication par plusieurs spécialistes du

Le but de la science ouverte, c'est aussi d'améliorer ce processus de révision par les pairs. Cela permet à plus de gens, dotés de compétences diverses, d'estimer la qualité des données, la pertinence des analyses, leur interprétation et donc la validité des conclusions de l'article.

domaine, ne permet-il pas de s'assurer que l'article est solide?

DR La révision par les pairs est une étape importante de la recherche et un sujet qui suscite beaucoup de discussions en ce moment. Elle contribue à empêcher la publication d'études erronées, mais ce n'est pas parfait. On le voit d'ailleurs au nombre grandissant d'articles qui sont rétractés.

Un des problèmes est que les éditeurs et les réviseurs sont eux aussi des chercheurs avec des horaires surchargés. Ils n'ont souvent pas le temps de revoir en profondeur les analyses présentées dans une étude et encore moins les données, quand elles sont disponibles.

Certains journaux, comme *Science* ou *Nature*, exigent depuis quelques années que les données nécessaires pour reproduire les analyses statistiques, figures et tableaux dans un article soient disponibles au moment de la publication. D'autres journaux se limitent à le recommander – sans obligation – alors que d'autres encore n'ont pas de politique à cet effet.

Dans tous les cas, la qualité des données partagées est très variable, donc il y a encore place à l'amélioration. Avec des collègues, nous avons établi un barème pour évaluer cet aspect – plus précisément la complétude et la réutilisabilité des données. Une bonne pratique est de partager les données brutes dans un format qui facilite leur réutilisation, comme un fichier texte ou CSV. Pour l'instant, c'est rarement le cas. Ce n'est parfois qu'un PDF ou la photo d'un tableau Excel. C'est inexploitable!

Le but de la science ouverte, c'est aussi d'améliorer ce processus de révision. Cela permet à plus de gens, dotés de compétences diverses, d'estimer la qualité des données, la pertinence des analyses, leur interprétation et donc la validité des conclusions de l'article. Avec plus de transparence, on augmente la confiance.

QS N'y a-t-il pas de limites éthiques à rendre publiques des données sensibles ou confidentielles?

DR Ce sont des aspects importants, mais il est possible de prendre des précautions. C'est ce que fait le Neuro [Institut-hôpital neurologique de Montréal], qui fait partie des établissements avant-gardistes. Avec d'autres instituts de neurosciences, il

a travaillé pour s'assurer de rendre les données publiques sans porter atteinte à la confidentialité des dossiers des patients.

On peut ainsi anonymiser les résultats ou partager les métadonnées en disant qu'une étude a été menée auprès de 2 000 patients sur telle maladie. Les personnes ou organismes qui veulent avoir accès aux détails doivent présenter leur demande à un comité qui veille à l'utilisation appropriée des données.

QS Dans certains domaines, comme l'astronomie ou la physique, les données sont publiques. Pourquoi est-ce différent en biologie?

DR Dans certains secteurs où les expériences et les outils coûtent extrêmement cher, comme l'accélérateur du CERN [Organisation européenne pour la recherche nucléaire] ou les grands télescopes, il y a tellement de données que ce serait impossible pour une seule équipe de les analyser. On les rend accessibles pour en tirer un maximum d'informations. En génomique aussi, la culture du partage est plus ancienne.

En revanche, dans les domaines où les expériences sont plus petites, comme en écologie et en biologie de l'évolution, ce n'est pas répandu. Pour les gens qui effectuent des études de terrain, qui travaillent d'arrache-pied pour obtenir du financement et qui suivent leurs populations d'oiseaux pendant des décennies par exemple, le partage des données est moins attirant. Ils ont peur que d'autres scientifiques publient des articles avec leurs données qu'ils auraient pu publier eux-mêmes.

QS Quand on sait qu'il y a une véritable course à la publication, on comprend ces réticences!

DR En effet, c'est un tiraillement pour les chercheurs : d'un côté, ils peuvent faire avancer la science en partageant leur travail ; de l'autre, ils craignent de favoriser leurs concurrents et de nuire à leur carrière. Et l'on sait qu'en ce moment, c'est dur de faire carrière en recherche, la compétition est féroce. Mais les avantages sont indéniables : d'abord, cela oblige les scientifiques à mieux gérer et organiser leurs données, pour que d'autres puissent les comprendre. Eux-mêmes y gagnent : ils peuvent également mieux les exploiter, même quelques années plus tard. Les chercheurs ne sont pas très doués sur ce plan :

leurs données sont souvent consignées dans une multitude de fichiers Excel avec des codes de couleurs mal définis ! Au bout du compte, c'est donc un gain de productivité. Et souvent, le partage peut mener à des collaborations lorsque le jeu de données intéresse une autre équipe. Mais il faut que les chercheurs soient récompensés pour cette contribution scientifique.

Une étudiante et moi avons interrogé 140 chercheurs de 20 universités canadiennes, et l'on a constaté que la majorité d'entre eux estiment que les avantages du partage sont plus importants que les coûts. Seulement 20 % des chercheurs rapportent un aspect négatif, déplorant le plus souvent le temps que cela leur prend. Certains affirment que leurs données ont été mal interprétées par d'autres ou qu'ils se sont fait couper l'herbe sous le pied pour un article, mais c'est très rare.

QS Quelle forme pourraient prendre les récompenses aux chercheurs qui donnent dans la science ouverte ?

DR Il faut que les organismes de financement et les universités encouragent les chercheurs à travailler de façon collective et accordent de l'importance au fait de partager des données de qualité, qui sont réutilisées par d'autres. Le nombre de publications ne doit pas être le seul critère pour une embauche ou l'obtention d'un financement.

Il y a un mouvement dans cette direction. Je pense entre autres à la Déclaration de San Francisco sur l'évaluation de la recherche, signée par plusieurs universités, qui permet d'évaluer les chercheurs non seulement sur la publication d'articles, mais aussi sur tous les autres produits de la recherche [prépublication d'articles, jeux de données, logiciels ouverts, retombées sociales, etc.]. On veut favoriser un équi-

libre entre les chercheurs qui recueillent des données et ceux qui les analysent, car il ne faut pas non plus dévaloriser la collecte de données, qui est fastidieuse et qui ne se traduit pas toujours par un grand nombre de publications.

QS Vous proposez un vrai changement de paradigme alors !

DR Oui, c'est une révolution. Il y a un appétit de la part des jeunes chercheurs pour changer le système. C'est possible et cela a déjà commencé. Selon moi, le partage des données pourra démocratiser la recherche. Les pays plus riches peuvent se payer de gros laboratoires, des équipements, des voyages... Si les données sont ouvertes, des chercheurs avec moins de moyens y auront accès aussi. Cela ouvre la voie à un rééquilibrage des ressources, y compris au sein d'un même pays. ●

*Comment
SAPIENS
a-t-il cessé d'être
UNE BÊTE
comme les autres ?*

*Une formidable synthèse
des plus récentes découvertes
sur l'évolution.*

ÉDITIONS
MULTIMONDES  
editionsmultimondes.com

Également offert en version numérique 

