

Recherche reproductible : création d'un réseau national; « un enjeu de formation » (S. Cohen-Boulakia)



Journées de la recherche reproductible, les 08 et 09/03/2023 - © D.R.

« La reproductibilité est le fait d'être capable de refaire une expérience et d'obtenir le même résultat. (...) Les causes premières de non reproductibilité ne sont pas la fraude mais plutôt la pression à publier et le manque de formation », déclare [Sarah Cohen-Boulakia](#), professeure en informatique et directrice adjointe de l'institut DataIA à l'Université Paris-Saclay, et chercheuse au LISN (Laboratoire interdisciplinaire des sciences du numérique) (Université Paris-Saclay, CNRS (Centre national de la recherche scientifique)), à News Tank le 10/03/2023.

Elle s'exprime à la suite des journées de la recherche reproductible, qu'elle co-organisait à Paris les 08 et 09/03/2023. Cet événement visait à initier un réseau national interdisciplinaire. Le MESR (Ministère de l'enseignement supérieur et de la recherche) est « prêt à nous soutenir (...), le Coso (Comité pour la science ouverte) pourrait nous donner un amorçage de moyens. »

Elle s'exprime à la suite des journées de la recherche reproductible, qu'elle co-organisait à Paris les 08 et 09/03/2023. Cet événement visait à initier un réseau national interdisciplinaire. « Ce réseau doit nous permettre d'être plus visible et que la France ait son mot à dire et son rôle au niveau de la coordination des différents réseaux sur la reproductibilité. »

Côté formation, l'enjeu pour le futur réseau concerne à la fois :

- « la formation initiale : nos étudiants lorsqu'ils sortent à bac +3, bac +5 ou bac +8 doivent avoir une idée de ce que veut dire travailler de façon reproductible ;
- la formation continue, pour ceux qui sont déjà diplômés, parce que la reproductibilité fait partie des bonnes pratiques que toutes et tous doivent mettre en place ».

Sarah Cohen-Boulakia pointe aussi « un enjeu plus politique », en lien avec l'évaluation de la recherche : « Au rythme effréné auquel on nous demande aujourd'hui de publier et d'avoir des résultats coûte que coûte, il ne faut pas s'étonner que tous les scientifiques ne prennent pas le temps de vérifier que leur recherche est bien reproductible. »

« Nous souhaiterions inciter les agences de financement nationales et européennes à demander aux chercheurs financés un engagement à faire une science reproductible. »

« Nous sommes en train de construire un réseau national »

Les 08 et 09/03/2023, vous co-organisez des « journées de la recherche reproductible » visant à faire un état des lieux de la reproductibilité en France, grâce au soutien du GDR Madics et de l'Institut Pasteur à Paris. Pourquoi ?

Nous sommes en train de construire un réseau national de recherche reproductible, qui n'a pas encore de statut. Néanmoins, cela fait une bonne dizaine d'années que nous sommes très nombreux à travailler sur l'animation de communauté autour de la recherche reproductible.

J'étais auparavant co-responsable avec Christophe Blanchet du groupe de travail sur la reproductibilité au sein du GDR Madics (Masses de données, informations et connaissances en sciences) du CNRS et de l'Institut français de bioinformatique, dans lequel Frédéric Lemoine (ingénieur de recherche en bio-informatique à l'Institut Pasteur à Paris) a joué un rôle clé. Ce groupe rassemblait entre 150 et 200 personnes selon les événements. Ce n'est pas rien, 200 personnes prêtes à se déplacer régulièrement, à faire des retours d'expériences et à se poser des questions de reproductibilité. Aujourd'hui la communauté bioinformatique dont je fais partie a bien compris l'enjeu de la reproductibilité.

« Des réseaux de reproductibilité un peu partout dans le monde »

Des réseaux de reproductibilité sont nés un peu partout dans le monde, en Europe et ailleurs, et nous avons été nombreux à avoir été contactés indépendamment, à peu près dans le même pas de temps, par des collègues nous incitant à faire quelque chose pour que la France monte elle aussi son réseau. Selon les pays, ces réseaux sont poussés par différentes disciplines comme la psychologie pour nombre d'entre eux. En France, nous cherchons à impliquer les collègues largement en invitant toutes les disciplines depuis l'informatique jusqu'à la physique et en passant par la médecine ou l'histoire.

Nous nous sommes contactés et avons décidé d'organiser ces journées, avec Arnaud Legrand, directeur de recherche CNRS en informatique, spécialiste de la reproductibilité des calculs à Grenoble, et Nicolas Rougier, directeur de recherche Inria (Institut national de recherche en informatique et en automatique) dans le domaine des neurosciences à Bordeaux. Nous avons également associé à l'organisation Frédéric Lemoine.

Notre volonté était d'ouvrir cet événement et d'en faire un appel large à la communauté, car nous sommes très centrés sur la reproductibilité computationnelle. Or, il y a beaucoup d'autres formes de reproductibilité selon les disciplines.

« La reproductibilité est un domaine de recherche en soi »

Selon Sarah Cohen-Boulakia, « la reproductibilité, c'est le fait d'être capable de refaire une expérience et d'obtenir le même résultat. Il faut avoir conscience que reproduire des conditions expérimentales est difficile, l'environnement n'est pas toujours maîtrisé jusqu'au bout, il est donc important de savoir tout ce qui joue dans l'obtention d'un résultat.

C'est assez intuitif en chimie par exemple, nous l'avons tous vécu à l'école, il faut connaître notamment la température ou le PH d'une expérience pour qu'elle se reproduise. En revanche, c'est moins attendu dans d'autres disciplines, comme l'informatique. Une machine on s'attend à ce qu'elle donne toujours le même résultat. Or, dans un calcul complexe, beaucoup d'éléments peuvent jouer dans un résultat, et il est important de savoir lesquels pour expliquer pourquoi ces calculs sont reproductibles ou non ».

Interrogée sur les raisons pour lesquelles elle s'est investie sur ce sujet, Sarah Cohen-Boulakia indique : « La reproductibilité, j'y travaille depuis toujours. En tant que scientifique, comprendre les bonnes pratiques qui permettent de faire de la science cumulative, c'est-à-dire de pouvoir se reposer sur des résultats précédents, est fondamental et repose sur la reproductibilité des résultats. Pour cela, il faut connaître les méthodes et données que l'on génère, garder la trace des raisons pour lesquelles elles sont comme ça et pas autrement, ne pas oublier les paramètres choisis dans un calcul, etc. Toutes ces informations que l'on trace, il faut leur donner du sens, les organiser, les comprendre : la reproductibilité est un domaine de recherche en soi. »

Le premier objectif de ces journées était de « poser un regard interdisciplinaire » sur les approches et définitions respectives des chercheurs et de « dresser un premier état des lieux de la reproductibilité en France ». Cet objectif a-t-il été atteint ?

En effet, chacun avait une idée de la problématique de la reproductibilité dans sa propre discipline. Ces journées visaient à faire se confronter des points de vue : que signifie être reproductible en médecine par exemple ? Ce n'est pas la même chose qu'en droit, en informatique, en histoire des sciences, etc. Nous avons ainsi demandé à chacun de présenter en dix minutes ce que voulait dire être reproductible dans son domaine ; quelles sont les bonnes pratiques existantes, etc. Cela a été absolument passionnant !

Nous nous sommes posé des questions de vocabulaire pour essayer de se comprendre, et des questions d'enjeux en particulier de formation.

Nous nous sommes aperçus, en matière de vocabulaire, que nous avons tous à peu près le même jargon, mais avec des nuances : par exemple, la gradation peut être inversée entre les termes « répliquable » et « reproduire », qui transmettent tous deux l'idée de refaire quelque chose et d'avoir les mêmes résultats. Nous avons réussi à collecter des définitions de la reproductibilité dans différentes disciplines et nous pourrions mettre en place de petits groupes de travail pour élaborer un lexique.

« Des définitions de la reproductibilité dans différentes disciplines »

Quels sont les enjeux qui se posent pour le futur réseau ?

Nous avons un énorme enjeu de formation, qu'elle se place au niveau :

- de la formation initiale : nos étudiants lorsqu'ils sortent à bac +3, bac +5 ou bac +8 doivent avoir une idée de ce que veut dire travailler de façon reproductible, cela nous semble fondamental ;
- de la formation continue : pour ceux qui sont déjà diplômés, parce que la reproductibilité fait partie des bonnes pratiques que toutes et tous doivent mettre en place.

C'est probablement l'un des enjeux les plus importants pour notre futur réseau. Une prise de conscience est nécessaire, les collègues comme les étudiants ne se rendent pas compte de l'enjeu, que changer une toute petite chose dans une expérience peut avoir un impact. Cela implique de les confronter au problème.

Ainsi, dans les formations initiales, former à la reproductibilité signifiera mettre les étudiants face à des situations où ils réalisent que ce n'est pas si facile de rendre ses résultats reproductibles. Une fois cette prise de conscience effectuée, il restera à former des formateurs ! Tout cela s'inscrit dans la même veine que la science ouverte, le partage des données et des méthodes.

« On va demander aux différents collègues de ralentir »

Il y a aussi un enjeu plus politique, probablement sur le long terme : face à l'injonction à produire, au rythme effréné auquel on nous demande aujourd'hui de publier et d'avoir des résultats coûte que coûte, il ne faut pas s'étonner que tous les scientifiques ne prennent pas le temps de vérifier que leur recherche est bien reproductible car cela a un coût en temps et ralentit la recherche. Or, la science reproductible est la base de la science cumulative. En étant reproductibles, nous avons conscience que l'on va demander aux différents collègues

de ralentir. Cela va dans le même sens que la communauté travaillant sur la réforme de l'évaluation de la recherche.

Le second objectif de ces journées était de réfléchir à la définition et à la gouvernance du futur réseau national, « permettant d'échanger des informations sur les formations, les ateliers, les conférences et de pérenniser l'organisation de ces journées », et son positionnement vis-à-vis d'instances comme le Coso (Comité pour la science ouverte).

Isabelle Blanc, administratrice ministérielle des données, des algorithmes et des codes sources au MESR (Ministère de l'enseignement supérieur et de la recherche), a présenté le Coso (Comité pour la science ouverte) et indiqué que le MESR était prêt à nous soutenir en tant que réseau. Nous ne savons pas encore sous quelle forme, mais le Coso pourrait nous donner un amorçage de moyens pour nous aider à nous structurer et pour nous offrir un soutien logistique.

Ce réseau doit nous permettre d'être plus visible et que la France ait son mot à dire et son rôle au niveau de la coordination des différents réseaux sur la reproductibilité. L'idée est de rentrer dans un réseau de réseaux internationaux pour échanger scientifiquement sur les pratiques, les outils et les conseils et formations qu'on pourrait donner aux collègues.

« Le Coso pourrait nous donner un amorçage de moyens »

Nous souhaiterions aussi inciter les agences de financement nationales et européennes à demander aux chercheurs financés un engagement à faire une science reproductible, comme pour la science ouverte. Cela commence à se faire au travers des plans de données, puisqu'il faut indiquer comment ces données ont été obtenues et les décrire.

Le futur réseau développera-t-il des liens avec l'Ofis (Office français de l'intégrité scientifique) ?

L'Ofis s'intéresse à l'intégrité scientifique. Il y a donc clairement des liens avec nos réflexions autour de la recherche reproductible, bien que les sujets autour de la reproductibilité ne se limitent pas à des problèmes d'intégrité.

« Pression à publier et manque de formation : causes premières de non reproductibilité »

Comme Nicolas Rougier l'a rappelé lors de l'introduction de ces journées, une enquête réalisée par Nature en 2016 auprès de 1 500 chercheurs montre que les causes premières de non reproductibilité ne sont pas la fraude mais plutôt la pression à publier et le manque de formation.

Si nous souhaitons pouvoir interagir avec l'Ofis, notre positionnement restera différent et ne s'attaquera pas aux mêmes problématiques.

Quelles sont les prochaines étapes ?

Nicolas Rougier et plusieurs volontaires sont en train de finaliser un rapport de quelques pages résumant ce qui s'est passé durant ces deux journées. Nous avons aussi commencé à définir certains groupes de travail et avec Frédéric Lemoine et Arnaud Legrand nous allons envoyer un email à toutes les personnes inscrites aux journées en leur demandant si elles souhaiteraient se positionner pour contribuer à structurer le réseau.

D'ici avril, nous collecterons les résultats et organiserons une réunion en visio avec les personnes intéressées pour définir plus précisément les contours des groupes et sous-groupes de travail, sur les pratiques, les formations, etc., ainsi que notre « plan d'attaque » !

De nouvelles journées nationales seront organisées à l'automne, avec l'objectif d'en avoir une à deux par an à l'avenir et en essayant d'attirer un peu plus de scientifiques dans les domaines de la reproductibilité observationnelle, des sciences humaines et sociales : histoire, géographie, sociologie, droit, mais aussi nous l'espérons ... journalisme !

« De nouvelles journées nationales à l'automne »



Sarah Cohen-Boulakia

[Directrice adjointe de l'institut DataIA @ Université Paris-Saclay \(EPE\)](#)

[Professeure des universités et chercheuse au LISN \(Laboratoire interdisciplinaire des sciences du numérique - CNRS, Université Paris-Saclay\) @ Université Paris-Saclay \(EPE\)](#)

[Directrice du GDR Madics \(Masses de données, informations et connaissances en sciences\) @ Centre national de la recherche scientifique \(CNRS\)](#)

Parcours

Depuis septembre 2022	Université Paris-Saclay (EPE) Directrice adjointe de l'institut DataIA
Depuis janvier 2021	Université Paris-Saclay (EPE) Professeure des universités et chercheuse au LISN (Laboratoire interdisciplinaire des sciences du numérique - CNRS, Université Paris-Saclay)
Depuis janvier 2020	Centre national de la recherche scientifique (CNRS) Directrice du GDR Madics (Masses de données, informations et connaissances en sciences)
Septembre 2017 - septembre 2020	Université Paris-Sud (Paris 11) Professeure des universités et chercheuse au LRI (Laboratoire de recherche en informatique - CNRS, Université Paris-Sud)
Septembre 2007 - septembre 2017	Université Paris-Sud (Paris 11) Maîtresse de conférences et chercheuse au LRI (Laboratoire de recherche en informatique - CNRS, Université Paris-Sud)
Septembre 2014 - mars 2016	Institut national de recherche en informatique et en automatique (Inria) Délégation dans le cadre de l'IBC (Institut de biologie computationnelle)
Décembre 2005 - août 2007	Université de Pennsylvanie (États-Unis) Post-doctorat

Établissement & diplôme

2015 - 2015	Université Paris-Sud (Paris 11) Habilitation à diriger des recherches
2005 - 2005	Université Paris-Sud (Paris 11) Thèse de doctorat
2001 - 2002	Université Paris-Sud (Paris 11) DEA en informatique

Fiche n° 48802, créée le 15/03/2023 à 10:26 - Màj le 15/03/2023 à 10:48