
Pourquoi les données de recherche ne sont-elles pas publiées ?

Why are Research Data not Published?

Ben Kaden



Édition électronique

URL : <http://journals.openedition.org/edc/8783>

DOI : 10.4000/edc.8783

ISSN : 2101-0366

Éditeur

Université de Lille

Édition imprimée

Date de publication : 1 juin 2019

Pagination : 137-146

ISBN : 978-2-917562-21-5

ISSN : 1270-6841

Référence électronique

Ben Kaden, « Pourquoi les données de recherche ne sont-elles pas publiées ? », *Études de communication* [En ligne], 52 | 2019, mis en ligne le 01 janvier 2021, consulté le 21 janvier 2021. URL : <http://journals.openedition.org/edc/8783> ; DOI : <https://doi.org/10.4000/edc.8783>

© Tous droits réservés

Pourquoi les données de recherche
ne sont-elles pas publiées ?

Why are Research Data not Published?

Ben Kaden

Humboldt-Universität zu Berlin
ben.kaden@ub.hu-berlin.de

Résumé / Abstract

L'article décrit les raisons pour lesquelles des chercheurs ne partagent pas leurs données de recherche. À partir d'une session lors d'un *BarCamp* de la Communauté Scientifique *Leibniz* et des résultats du projet *eDissPlus* sur les données de recherche liées aux thèses de doctorat, l'auteur regroupe et commente les explications et arguments pour ne pas déposer et/ou ouvrir les données de recherche dans un entrepôt de données. Il s'agit de raisons techniques ou juridiques mais aussi d'attitudes personnelles ou d'absences de ressources appropriées. Trois enjeux sont identifiés : la disproportion entre l'effort et le bénéfice, une situation juridique peu claire et un manque d'infrastructures opérationnelles. Trouver des solutions nécessite une médiation active et un dialogue permanent entre les professionnels de l'information et les communautés scientifiques.

Mots-clés : science ouverte, données de recherche, accès libre, partage de données, comportement données.

This article describes the reasons why researchers do not share their research data. Based upon a session at a BarCamp of the Leibniz Association and the results of the eDissPlus project on research data related to doctoral theses, the author summarises and analyses the explanations and arguments given for not depositing and/or not opening research data in a data repository. The reasons are technical or legal, but also include personal attitudes or lack of appropriate resources. Three issues are identified: the disproportion between effort and benefit; an unclear legal situation; and lack of operational infrastructure. Finding solutions requires active mediation and ongoing dialogue between information professionals and scientific communities.

Keywords: open science, research data, open access, data sharing, data behaviour.

¹[Note préliminaire : Une première version de ce texte a été publiée le 13 mars 2018 dans le blog *LIBREAS. Library Ideas*². Étant donné que la (non) publication des données de recherche paraît l'un des sujets d'actualité de la publication scientifique, ce texte est republié ici dans la rubrique « Documentation » de la revue *LIBREAS* sous une forme révisée. L'auteur travaille actuellement sur le thème de la publication des résultats scientifiques liés aux thèses de doctorat dans le cadre du projet *eDissPlus* de la DFG, qui constitue le contexte immédiat de ce texte].

Une question majeure et probablement encore trop peu systématisée dans toutes les discussions sur la science ouverte, du moins pour les infrastructures actives dans ce domaine, est : « Qu'est-ce qui s'oppose réellement à la publication des données de la recherche ? ». En tout état de cause, l'expérience de l'*Open Access* et plus encore de l'*Open Science* ou de l'*Open Scholarship* montrent qu'il existe souvent un écart considérable entre les souhaits, les objectifs et les idées de l'infrastructure de recherche et les représentants scientifiques particulièrement engagés dans ce domaine et un autre groupe qu'on appellera ici par un raccourci *mainstream* de la science.

Force est de constater que l'un des enseignements importants, quoique pas tout à fait surprenant, des années de débat sur la science ouverte est que la plupart des chercheurs veulent faire de la recherche et ceci avant tout dans une logique de culture de publication scientifique qui leur est familière. Les lacunes des systèmes de publication sont tout à fait reconnues, mais ne sont traitées plus en profondeur que si elles entraînent des obstacles notables à la recherche individuelle. Dans la plupart des cas, cependant, les chercheurs ne veulent pas devenir des innovateurs en matière de communication scientifique et de solutions d'infrastructure d'une manière qui conduirait à un report de l'attention de l'objet de recherche vers ces métastructures de communication scientifique. Ainsi, tant que la détresse des chercheurs due aux systèmes et pratiques existants n'est pas excessivement élevée et dans la mesure où les formes de publication traditionnelles jouent toujours en faveur de leur carrière, même des *keynotes* très engagés et sophistiqués en faveur de la science ouverte ne changeront pas grand-chose. Il est donc essentiel que les bibliothèques scientifiques et les autres acteurs de l'infrastructure scientifique connaissent les attentes, les défis et les objectifs des différentes communautés. Cela comprend également les raisons pour lesquelles les données et matériels de recherche sont rendus plus ou moins accessibles selon les différentes disciplines, mais

1 Cet article est la version traduite de Kaden B. (2018). « Warum Forschungsdaten nicht publiziert werden ». In *LIBREAS. Library Ideas*, vol. 33, s.p. Disponible sur <https://libreas.eu/ausgabe33/kaden-daten/> (page consultée le 2 mars 2019).

2 <https://libreas.wordpress.com/2018/03/13/forschungsdatenpublikationen/>.

d'une manière générale et par rapport à l'idéal de la science ouverte, finalement très peu.

Lors de l'*Open Science Bar-camp*³ du réseau de recherche *Leibniz Science 2.0* qui s'est tenu à Wikimedia le 12 mars 2018, il y a eu une session intitulée « *Valid reasons for opting out of sharing openly*⁴ ». Conforme aux principes de la science ouverte, certains points clés de cette session ont été mis à disposition dans un *Etherpad* pour tous ceux qui n'ont pas pu y participer⁵.

J'ai pris la liberté de regrouper et de reformuler ces points clés. Suite à cette liste, j'ai ajouté quelques points clés du projet *eDissPlus*⁶, qui portait sur l'accessibilité des données de recherche liées aux thèses de doctorat. D'un point de vue général, il s'agit également de données qui, notamment en sciences humaines et sociales, seraient plutôt décrites comme des matériels de recherche, par exemple des textes sources ou des images.

Investissement

- Les chercheurs préfèrent investir leur temps dans la recherche elle-même plutôt que dans l'organisation d'un processus d'échange de données de recherche.
- La planification du projet ne comprend pas le temps ou les ressources humaines nécessaires à la préparation des données de recherche à des fins de partage ou de publication.
- La publication ou l'accessibilité des données de recherche n'a pas été prise en compte lors de la planification du projet ou de la préparation du plan de gestion des données de recherche et est trop longue à mettre en œuvre par la suite⁷.

Loi sur la protection des données

- La publication ou la mise à disposition de données de recherche est exclue pour des raisons de protection des données.
- Il n'y a pas de consentement éclairé pour la divulgation ou la publication de données personnelles.

Équipements institutionnels / infrastructures

- Sa propre institution n'offre pas un soutien suffisant en termes d'in-

3 <http://www.open-science-conference.eu/barcamp/>.

4 « Raisons valables de refuser de partager ouvertement » (traduction par la rédaction).

5 https://etherpad.wikimedia.org/p/oscibar2018_session13.

6 <https://www2.hu-berlin.de/edissplus/>.

7 Au sujet des plans de gestion des données de recherche, comparez en détail avec : Kaden B. (2018). « Forschungsdatenmanagementpläne sind eine Grundbedingung guter Wissenschaft. Meint Nature ». In *LIBREAS.Tumblr*, 21.03. Disponible sur <http://libreas.tumblr.com/post/172103346561/> (page consultée le 2 mars 2019).

frastructure et de conseils pour rendre les données de recherche disponibles ou les publier.

Conditions institutionnelles

- Le règlement de l'épreuve interdit aux doctorants de publier des parties de leur thèse de doctorat avant la fin de la procédure de doctorat.
- Il n'y a pas de critères de sélection formels pour déterminer quelles données de recherche devraient être rendues accessibles et comment.

Possibilités et compétences de partage / publication

- Les scientifiques ne savent pas où ils peuvent déposer leurs données à des fins de divulgation.
- Les scientifiques ne sont pas suffisamment formés pour rendre les données de recherche accessibles ou les publier conformément aux normes de publication scientifique.
- Les publications de données de recherche devraient être soumises à une procédure d'évaluation par les pairs (*peer review*), qui n'existe peut-être pas encore. La publication de données de recherche sans *peer review* est rejetée.

Attitude personnelle / contrôle des données / éthique de la science

- Les scientifiques ne sont pas intéressés par le thème de la science ouverte.
- Les scientifiques aimeraient savoir qui utilise leurs données de recherche, c'est pourquoi ils ne les transmettent que sur demande personnelle ou se réservent le droit de refuser de les transmettre.
- Les partenaires dans un projet de recherche s'opposent à la mise à disposition ou à la publication des données de recherche générées dans le cadre du projet.
- La divulgation des données de recherche est délibérément refusée parce que de telles suggestions et spécifications sont interprétées comme une atteinte à la liberté académique personnelle.
- Les données de recherche sont considérées comme n'étant pas suffisamment pertinentes pour être transmises.
- Les scientifiques veulent empêcher que leurs données de recherche ne soient réutilisées à des fins qu'ils ne veulent pas.
- Des doutes subsistent quant à la capacité de tiers d'utiliser les données de recherche ou les matériels de recherche conformément aux standards scientifiques.
- On craint qu'en rendant les données de recherche accessibles, les faiblesses dans la collecte et l'analyse des données ne deviennent visibles.
- Les données concrètes de la recherche sont manipulées d'une manière qui doit rester cachée.

Droits d'édition, d'auteur et d'utilisation

- Les scientifiques ont transféré les droits d'utilisation et d'exploitation à un éditeur scientifique en vertu d'un contrat de droits d'éditeur et n'ont donc aucun moyen de partager et de publier des données de recherche.
- Les doctorants dont les projets de recherche sont menés en collaboration avec des tiers n'ont que des droits de disposition limités sur leurs données de recherche. Cela vaut en particulier pour les collaborations avec des partenaires commerciaux.
- On ne sait pas qui est le propriétaire légal des données de recherche respectives.

Autres domaines du droit / éthique des sciences

- Le sujet de recherche est trop sensible pour que les documents et les données de recherche soient librement accessibles à l'échelle internationale.
- On ne sait pas exactement qui assumera la responsabilité à long terme des données et des documents de recherche respectifs.

Sociologie de la science

- Les données et les matériels de recherche sont considérés comme du capital scientifique et sont (encore) retenus pour une exploitation dans un projet futur.
- Les données et les matériels de recherche doivent être cités comme un atout exclusif pour une demande de financement de projet. S'ils sont disponibles gratuitement, on a l'impression que les chances de financement diminuent.
- Dans un premier temps, les données de recherche doivent faire l'objet de recherches plus approfondies sur une base exclusive, c'est pourquoi la publication ou la mise à disposition est au mieux possible après un embargo.
- La publication ou le partage des données de recherche n'est pas suffisamment apprécié comme une performance scientifique.

Liberté académique

- Le principe de la science ouverte ne devrait pas être perçu comme une contrainte – au sens de la liberté académique, les chercheurs devraient décider par eux-mêmes s'ils veulent rendre leurs données de recherche accessibles, comment et quand.

Sur la base de l'expérience du projet *eDissPlus*, qui a examiné les attitudes des doctorants à l'égard de la publication des données de recherche, il est possible, comme annoncé, d'identifier un certain nombre d'autres obstacles et de différencier davantage les aspects mentionnés. Ce sont entre autres :

Investissement

- Comme nous l'avons déjà mentionné dans la planification de la recherche, l'effort nécessaire à la publication des données de recherche accompagnant la thèse est très rarement pris en compte dans la planification de la thèse et – si disponible – dans les plans de gestion des données de recherche.

Conditions institutionnelles et disciplinaires

- Dans de nombreux domaines, il y a un manque de normes pour la gestion des données de recherche et la publication de données de recherche qui pourraient fournir une orientation.
- Les politiques en matière de données de recherche sont souvent perçues comme inadaptées dans des cas individuels, notamment parce qu'elles ne tiennent aucunement compte, par exemple, des lois sur la protection des données ou d'autres restrictions légales à une éventuelle publication des données de recherche.
- Dans de nombreuses disciplines, il n'y a pas de débat approfondi et systématique sur l'importance et la forme de la publication des données de recherche dans le cadre de la communication scientifique.
- En règle générale, les règlements d'épreuves ne font aucune déclaration sur la publication des données de recherche et ne fournissent donc aucune orientation.
- Dans la plupart des cas, la publication des données de recherche ne sont pas requises pour l'obtention du doctorat.

Compétences et transfert de compétences

- La gestion des données de recherche est peu présente dans l'enseignement et la formation méthodologique. La question de la publication des données de recherche n'est généralement pas abordée du tout.
- Comme la publication des données de recherche n'est pas pratique courante dans de nombreuses disciplines, les doctorants n'entrent pas en contact avec ce genre de publications lorsqu'ils recherchent de la littérature. Ils ignorent donc souvent l'idée et la possibilité de publier des données de recherche.

Attitude personnelle / contrôle des données / éthique de la science

- Dans le contexte des projets de doctorat, les données de recherche sont souvent considérées davantage comme un moyen d'atteindre une fin que comme une variable distincte qui mérite d'être publiée.
- En raison du peu d'expérience avec la publication des données dans de nombreux domaines, les doctorants ne savent pas comment faire face et comment jouer le rôle de précurseur pour la culture de publication dans leur discipline, en plus de leur doctorat.

Aspects juridiques

- Les étudiants en doctorat sont souvent incapables d'évaluer correctement la publication des données de recherche sous l'aspect du droit d'auteur. Par exemple, l'application des licences *Creative Commons* semble souvent peu évidente pour des cas concrets.
- La loi sur la protection des données fait actuellement obstacle à la publication des données de recherche à caractère personnel liées à la thèse dans presque tous les cas de figure, notamment parce que la publication des données est rarement planifiée dès le départ et ne fait pas l'objet d'un consentement éclairé. L'obtention ultérieure d'une autorisation de publication n'est souvent pas possible ou est jugée trop fastidieuse.

Sociologie de la science / contrôle des données de recherche

- Dans la plupart des cas, la publication des données de recherche ne promet aucun gain de réputation supplémentaire. Dans certains cas, les évaluateurs considèrent même qu'une publication des données de recherche accompagnant une thèse est potentiellement nuisible.
- De nombreux doctorants considèrent les données de recherche comme un capital scientifique. S'il existe une volonté de communiquer les données de recherche, la mise à disposition sélective sur demande est clairement préférée à une diffusion globale sous forme d'une publication.

Ces deux listes ne sont certes pas exhaustives. De nombreux aspects mériteraient une analyse plus approfondie. Cependant, un tel classement montre aussi que même si les attitudes individuelles jouent un rôle important et que le lobbying pour le libre accès et la science ouverte a certainement un sens, elles porteront des fruits à plus long terme uniquement si des conditions externes favorables sont réunies, tant en termes d'infrastructures que de culture professionnelle.

Souvent, l'un des principaux obstacles est sans doute la disproportion entre l'effort et le bénéfice. En raison de l'absence de modèles de bonnes pratiques et de recommandations, la publication des données de recherche conforme aux normes scientifiques exige souvent une charge de travail supplémentaire relativement élevée, sans aucune récompense pour la carrière scientifique. Plus les infrastructures sont en mesure d'offrir des conseils et d'autres services facilement accessibles, mieux c'est. Bon nombre des doctorants interrogés dans le cadre du projet *eDissPlus* considéraient la bibliothèque universitaire comme un point de contact naturel pour toutes les questions relatives aux données de recherche, y compris pour les aspects qui relèvent davantage des structures de recherche et de la formation doctorale. Idéalement, la bibliothèque devrait offrir une gamme complète de services, allant des services de conseil et de *cloud* aux services d'archivage à long terme pour les structures de données complexes. Cependant, la question de savoir lequel de ces services peut être

offert, et comment, est un autre débat. Pour les doctorants compétents en matière de données, *GitHub* est un *benchmark*. Pour les autres, c'est plutôt *Dropbox*. Résultat : le besoin exprimé va dans les deux sens en même temps, avec des services de consultation complets ; et, soit dit en passant, ce besoin surestime clairement les capacités de développement que le secteur public peut fournir à ce point et à court terme.

Le deuxième grand défi réside dans une situation juridique peu claire en ce qui concerne les données de recherche. En janvier 2018, il y a eu un atelier sur ce sujet à l'Université européenne Viadrina de Francfort-sur-l'Oder⁸ qui, comme prévu, a apporté peu de réponses mais un aperçu encore plus approfondi de la complexité de la situation.

Enfin, dans de nombreuses disciplines, il y a un manque d'infrastructures opérationnelles, y compris de la part d'éditeurs ou d'autres prestataires commerciaux, pour diffuser les données de recherche d'une manière appropriée et durable. Le paradigme du *print*, toujours présent dans le format PDF, ne fonctionne plus du tout pour les données de recherche. Si l'on veut qu'elles fassent partie intégrante de la communication scientifique, il faut souvent tout d'abord trouver des médias et formats adéquats de présentation – un débat, soit dit en passant, qui se déroule dans de nombreux domaines, au mieux en marge. Implicitement, les enquêtes *eDissPlus* révèlent un autre obstacle majeur à la publication des données de recherche : dans certains cas, les données ne peuvent simplement pas être représentées avec les possibilités existantes.

Bien sûr, pour tous ces problèmes, un travail a été engagé pour trouver des solutions, même si la numérisation de la science, en particulier dans le secteur des infrastructures, pourrait être soutenue par les institutions et agences de financement de manière très différente de ce qui a été le cas jusqu'ici. Depuis cette année, par exemple, le serveur *edoc* a rendu possible la publication des données de recherche de la Humboldt-Universität de Berlin⁹. *Zenodo* peut également être considéré comme un exemple réussi de serveur de publication moderne pour une large gamme de résultats. Le fait que sur les serveurs de publication, les données de recherche soient indexées avec des métadonnées, qu'on leur attribue des DOI et qu'elles soient au moins partiellement cataloguées par les bibliothèques semble un premier pas vers la reconnaissance comme une véritable publication savante. Mais dans la plupart des cas, cela est déjà la limite de ce dont les bibliothèques et les infrastructures sont capables.

L'auto-organisation de la science oblige les communautés disciplinaires à se mettre d'accord sur la signification et la forme que la publication des données de recherche et d'autres matériels de recherche peut et doit prendre pour

8 http://www.forschungsdaten.org/index.php/Rechtliche_Aspekte_bei_digitalen_Forschungsdaten.

9 https://www.hu-berlin.de/de/pr/nachrichten/maerz-2018/nr_180309_00.

elles. Elles doivent discuter, examiner et décider elles-mêmes si, par exemple, elles veulent du *peer review*, si la publication des données de recherche complexes est prise en compte pour le recrutement, quels formats elles préfèrent et de quelles métadonnées elles ont besoin. Les infrastructures de leur côté peuvent montrer ce qui est possible, transmettre des expériences, des idées et des connaissances générales. Pour cela, nous avons besoin d'événements, tels que l'*Open Science Bar Camp*, et d'études sur la science (*science studies*), comme le projet *eDissPlus*. Sécuriser un niveau de connaissances en phase avec les pratiques et les souhaits des communautés d'une part, et les possibilités techniques d'autre part constituent un énorme défi en soi et en même temps une condition minimale pour tout développement d'infrastructure ciblée. Pour ce faire, il faut déjà des intermédiaires (« *Brückenakteure* ») qui connaissent aussi bien les communautés et publications scientifiques que les objectifs, les possibilités et les limites de l'infrastructure et de l'organisation de la recherche, et qui sont capables de faire le lien. Mais ces acteurs sont encore plus nécessaires lorsqu'il s'agit de franchir le pas d'une science numérique, à savoir relier l'infrastructure à la communication scientifique et, en certains points, directement à la recherche. Lors des *Bar Camps* et des ateliers, nous pouvons discuter longuement des raisons pour lesquelles les chercheurs ne publient pas leurs données. Cependant, des solutions tangibles et praticables ne peuvent en découler que quand ces discussions seront également menées avec les scientifiques eux-mêmes. Pour ce faire, il est nécessaire non seulement de connaître les deux parties, mais aussi de maintenir un dialogue permanent. Au début, j'ai remarqué que les chercheurs veulent d'abord faire de la recherche et s'occuper le moins possible des questions d'infrastructure. Cela change bien sûr dans le domaine des sciences numériques lorsque l'infrastructure et la recherche coïncident. Un bon exemple des difficultés de ce développement est celui des humanités numériques.

En tant que représentants de bibliothèques universitaires, par exemple, nous participons activement à des projets tels que *eDissPlus* pour comprendre la motivation et les objectifs des chercheurs comme groupe-cible. Cependant, à regarder de près, le concept même de groupe-cible pourrait devenir de plus en plus relatif pour se transformer progressivement en partenariat. Un premier *desideratum* serait dans l'immédiat un forum ou un cadre qui nous permettrait d'introduire des résultats tels que ceux présentés ci-dessus dans un dialogue global et constructif avec toutes les parties prenantes. Un autre souhait serait une structure stable et idéalement simple qui, après la fin de projets tels que *eDissPlus*, permettrait, au-delà de cette « anamnèse », de préciser où, comment et par quels moyens les verrous identifiés pourraient être levés. Cette situation est exemplaire de quelque chose de très général : les exigences d'une science numérique ouverte ne peuvent être réalisées qu'en tant que projet de l'ensemble du système scientifique.