

DATA PAPERS, DATA STATEMENTS

COMMENT VISIBILISER LE TRAVAIL
INVESTI DANS VOS DONNÉES

EN 15 MINUTES

Floriane Muller

12 mars 2024

DOI: [10.5281/zenodo.10806281](https://doi.org/10.5281/zenodo.10806281)



Illustration by [Freepik Storyset](#)

Bibliothèque



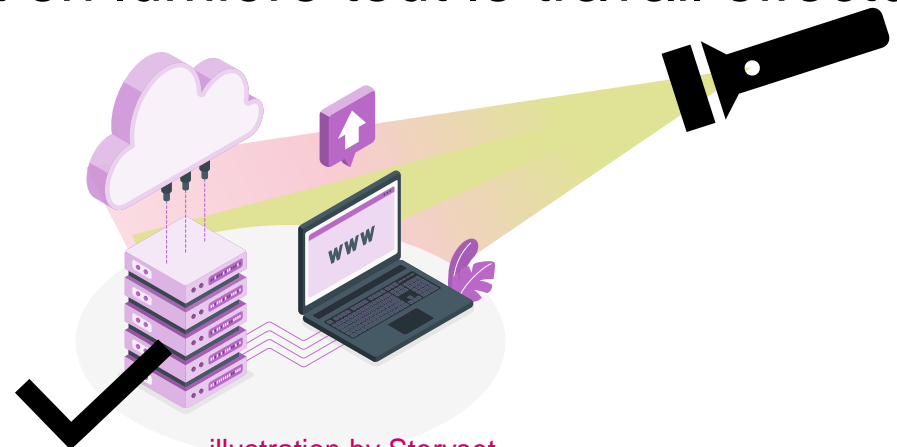
UNIVERSITÉ
DE GENÈVE

POURQUOI S'EN PRÉOCCUPER ?

- Les données sont précieuses. Vous y avez investi du temps et des ressources.
 - De nombreux bailleurs de fonds et organisations exigent désormais que les données soient déposées sur un ***data repository*** et rendues **FAIR***.
- *FAIR = Facile à trouver, Accessible, Interopérable, Réutilisable. “Aussi ouvert que possible, aussi fermé que nécessaire”.
- Déposer des données dans un data repository nécessite aussi des ressources (ex. pour décrire, documenter,...)
 - Améliorer leur visibilité met en lumière tout le travail effectué



[illustration by Storyset](#)



[illustration by Storyset](#)

COMMENT : 2 SOLUTIONS

→ Travailler sur la visibilité de vos données déposées dans le *repository*, même si ces dernières ne sont téléchargeables que sur demande.

Data papers



[Illustration by Storyset](#)

Data statements



[Illustration by Storyset](#)

DATA PAPERS

“A data paper is a journal publication whose primary purpose is to describe data, rather than to report a research investigation.” ([Chavan & Penev, 2011](#))



[illustration by Storyset](#)

Contient: Faits sur les données, p. ex.

- Liste des fichiers de données partagés et lien vers le *repository*
- Contexte de production (méthodes d'obtention, d'assurance de qualité, etc.)
- Couverture (temporelle et spatiale, critères d'exclusion, etc.)
- Limitations, et potentiel de réutilisation,
- Conditions de réutilisation
- ...

Ne contient pas:

- les hypothèses de recherche qui sont testées,
- la discussion, et les conclusions tirés des données.

[Lien vers un exemple pris au hasard](#)

UNE VARIÉTÉ DE NOMS ET DE REVUES

- Certaines revues (*data journals*) sont spécialisées dans les *data papers* et ne publient que ça.
- D'autres publient des *data papers* mais publient aussi d'autres types d'articles plus « classiques ».

Data article

Data descriptor

Data in brief

Data note

Data original article

Data paper

Database article

Database paper

Dataset paper

DATA JOURNALS: QUELQUES EXEMPLES

Titre	Éditeur	Type d'article	Indexé dans	OA	APC /coût	Modèle	Peer Review
BMC Research Notes	BioMed Central	Data Note	WoS PubMed	yes	1240 GBP	-	Blind peer review
Journal of Open Psychology Data	Ubiquity Press	Data Paper	-	yes	485 GBP	yes	oui
GigaScience	Oxford Univ. Press	Data Note	WoS PubMed	yes	1245 USD	-	Open Peer review
Research Data Journal for the Humanities and Social Sciences	Bill	Article	Scopus	yes	Pas d'APC	-	oui
Scientific Data	Nature	Data Descriptor	WoS PubMed	yes	2290 USD	yes	Blind peer review

Des listes de journaux publiant des *data papers* ou consacrés uniquement à ça existent en ligne, mais ne sont pas toujours à jour. Voir des exemples de listes, pluridisciplinaires [ici](#) ou [là](#), ou couvrant un domaine en particulier, comme par ex. la [biodiversité](#)

AVANTAGES ET INCONVÉNIENTS

D'UN *DATA PAPER*

- ✓ Augmente la **visibilité** des données déposées dans un *repository*
- ✓ Les données sont examinées par des pairs (**peer review**)
- ✓ Production scientifique, entièrement **citable et indexée** dans les bases des données académique → impact et reconnaissance de votre travail
- ✓ Décrit le jeu de données dans un **format facile à comprendre par des humains**, pour s'assurer qu'il est bien compris et réutilisé de manière appropriée
- ✓ Augmente le niveau de FAIRness des données
- ✗ rédiger, soumettre et publier le *data paper* prend du temps
- ✗ des APCs peuvent être dûs

DATA STATEMENTS

Quelques phrases incluses dans l'article de recherche publié fournissent des informations sur la disponibilité des données sur lesquelles l'article est basé.



Illustration by Storyset

Certaines revues ont commencé à offrir/exiger une section dédiée dans l'article : “**data availability statement**” ou “**data access statement**”. Généralement en 1^{ère} page ou à la fin de l'article.

But:

- **indiquer où et comment accéder aux jeux de données**, ou expliquer pourquoi ils ne peuvent pas être partagés librement.
- + fournir leur identifiant unique / lien pour les télécharger.**

EXEMPLES DE *DATA STATEMENTS*

Data Availability: Data are available on Yareta (DOI: [10.26037/yareta:d7oanpxk3jd2xhd7sg7qhicdv4](https://doi.org/10.26037/yareta:d7oanpxk3jd2xhd7sg7qhicdv4)) and code is available from GitLab at https://gitlab.unige.ch/KLab/mis_markers_misplacement.

(Fonseca et al. 2020)

Data Availability

The pseudonymized genotype and phenotype data, and the covariates used in the GWAS, are deposited in the institutional permanent archive of the University of Geneva (<https://doi.org/10/gq568g>), where all necessary information for data access can be found (the data will be made available upon signing a Data Use Agreement, assuring that the data will only be used in accordance with the restrictions of the informed consent, and agreeing that the data will be destroyed after the research is complete). All results needed to evaluate the conclusions in the paper are available in the main text or the **Supplementary Material** online, including **supplementary figs. S21 to S24**. (Mouterde et al. 2022)

De nombreux éditeurs fournissent des conseils concernant la formulation dans leurs directives de soumission d'auteur. Voir aussi les exemples de la bibliothèque de l'université de Manchester, couvrant la plupart des cas
<https://www.library.manchester.ac.uk/services/research/research-data-management/sharing/data-access-statements/>

Data availability

(Lavallard et al. 2023)

Underlying data

figshare: Formative assessments during COVID-19 pandemic: performance and experiences of medical students: <https://doi.org/10.6084/m9.figshare.21755882> (Lavallard et al., 2022)

This project contains the following underlying data:

- FormativeAssessmentCOVID.csv (this file contains different the raw quantitative data for variables such as score, duration of attempt, gender, year)
- SurveyCOVID.csv (this file contains the student's comments)

Please note that the responses to the qualitative survey questions are in French. Specific quotes or themes can be translated to English on request.

Extended data

figshare: Formative assessments during COVID-19 pandemic: performance and experiences of medical students: <https://doi.org/10.6084/m9.figshare.21755882> (Lavallard et al., 2022)

This project contains the following extended data:

- README Formative Assessment COVID.txt
- README Survey COVID.txt
- Supplementary data.png

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

AVANTAGES ET INCONVÉNIENTS D'UN *DATA STATEMENT*

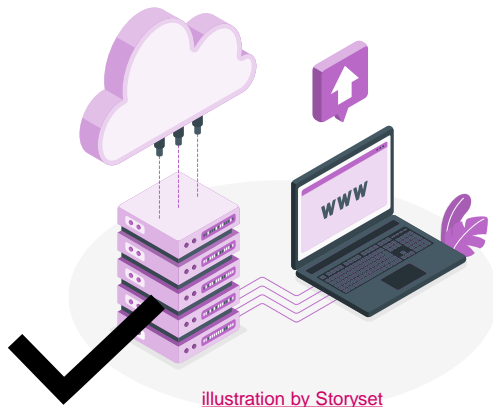
- ✓ Augmente la **visibilité** des données déposées dans un *repository*
- ✓ Montre les efforts que vous avez déployés et vos tentatives de **conformité aux exigences** de partage des données
- ✓ **Lie** votre article avec les données déposées **“de façon standardisée”** (analysable par les machines et facile à lire pour les humains)
- ✓ Permet de fournir plus d'informations qu'une simple citation de données (ex : conditions d'accès, données disponibles dans plusieurs *repositories*, etc.)
- ✓ Pas de frais supplémentaires pour vous
- ✗ Implique de connaître le lien direct/DOI au moment où vous soumettez votre article.



mais les *repositories* permettent maintenant souvent de
« réserver un DOI »

SCENARIO IDÉAL

Données déposées dans un *repository*



Obtiennent une bonne description et un identifiant/liens unique et persistant (DOI, accession number, etc.)

Ajouter un lien vers le(s) article(s) publié(s) dans la description du jeu de données sur le repository



illustration by Storyset

Un *data statement* est inclus dans la publication de scientifique exploitant les données, et il pointe vers le *data repository* où celles-ci sont déposées



illustration by Storyset

Lorsque vous déposez votre article dans le dépôt institutionnel de publications (Archive ouverte), indiquez également le lien vers les données



illustration by Storyset

Un *data paper* peut décrire les données (si aucune publication des résultats de recherche n'est faite, ou pour la compléter)

CONCLUSION

- Déposer des données de recherche dans un *data repository* prend du temps et demande du travail
- Il existe des solutions pour rendre ce travail visible (trouvable) et en obtenir le crédit.
- Si les données déposées mènent à un article, incluez un **data statement** dans ce dernier!
- Si aucune publication ne découle des données, ou si vous avez le temps, envisagez de publier un **data paper**
- Les références croisées améliorent la visibilité. **Faire des liens de renvoi partout où vous le pouvez en vaut la peine.**



QUESTIONS ?



**Equipe données de recherche
de l'Université de Genève**
researchdata-info@unige.ch

AUTRES FORMATIONS SUR LES DONNÉES DE RECHERCHE

15.03.2024

ATELIER

Deposit your research data in Yareta: getting started with the institutional tool

14:15 – 15:45, Online

19.03.2024

ATELIER

Gérer ses données selon les principes FAIR, en 15 minutes

09:00 – 09:15, En ligne, sur Zoom

ATELIER

Data Management Plan pour la CUREG: A quoi sert-il et comment le remplir?

14:15 – 15:45, En ligne

20.03.2024

ATELIER

Organiser et nommer ses fichiers efficacement, en 15 minutes

[agenda complet >>](#)



Floriane Muller, Bibliothèque de l'Université de Genève, 2024

This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

<https://creativecommons.org/licenses/by-sa/4.0/>

except for content where another author/license is mentioned.