



UNIVERSITÉ  
DE GENÈVE

BIBLIOTHÈQUE

# Data Papers, Data statements

## how to visibilize the work you've invested in your data

### in 15 minutes

Floriane Muller

February 14th, 2025

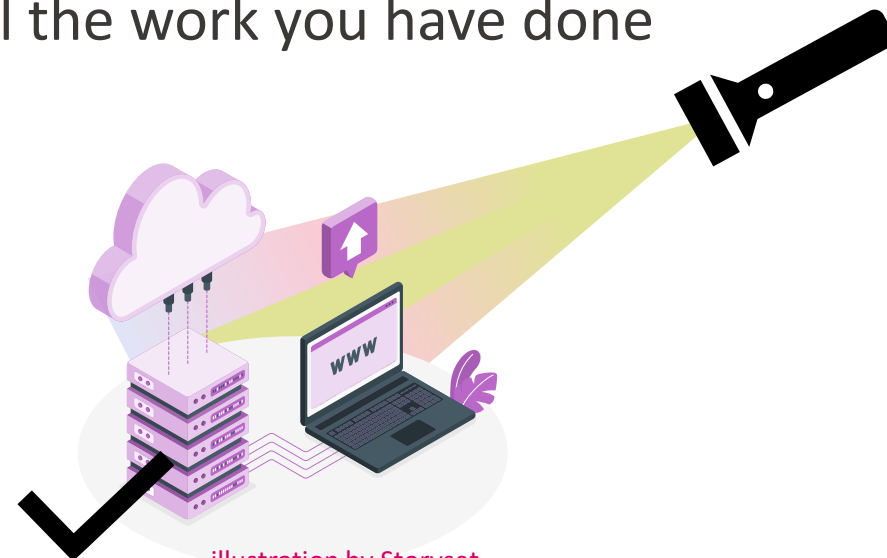
DOI: [10.5281/zenodo.14869430](https://doi.org/10.5281/zenodo.14869430)

## Why bother ?

- Data is precious. You have invested time and resources in it.
- Many funders and organisations now require that data should be uploaded to a **data repository** and made **FAIR\***.
  - \*FAIR = Findable, Accessible, Interoperable, Reusable. “As open as possible, as closed as necessary”.
- Uploading data to a repository takes resources too (e.g. describe & document, etc...),
- Enhancing its visibility sheds light on all the work you have done



[illustration by Storyset](#)



[illustration by Storyset](#)



## How: 2 solutions

→ Work on the visibility of your uploaded data, even if it is only accessible upon request.

### Data papers



[Illustration by Storyset](#)

### Data statements



[Illustration by Storyset](#)

# Data Papers

# 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

**Contains:** Facts about the data, e.g.

- ✓ List of the shared data files and link to the data repository
- ✓ Production context (methods by which it's been obtained, by which its quality has been ensured, etc.)
- ✓ Coverage (time and space, exclusion criteria, etc)
- ✓ Limitations, and reuse potential or conditions,
- ✓ ...

**Does not contain:**

- ✗ research hypothesis that are tested,
- ✗ arguments and conclusions drawn from the data.

[Link to a random example of a data paper](#)

## A variety of names and journals

- Some journals are specialized in data papers
- Other publish them alongside “regular articles”.

Data article

Data descriptor

Data in brief

Data note

Data original article

Data paper

Database article

Database paper

Dataset paper

## Data Journals: some examples

Title	Publisher	Article Type	Indexed in	OA	APC /Fees	Template	Peer Review
<a href="#">BMC Research Notes</a>	BioMed Central	Data Note	WoS PubMed	yes	1640 GBP	-	<a href="#">Blind peer review</a>
<a href="#">Journal of Open Psychology Data</a>	Ubiquity Press	Data Paper	-	yes	500 GBP	<a href="#">yes</a>	yes
<a href="#">GigaScience</a>	Oxford Univ. Press	Data Note	WoS PubMed	yes	2638 USD	-	<a href="#">Open Peer review</a>
<a href="#">Research Data Journal for the Humanities and Social Sciences</a> (discontinued ☹)	Bill	Article	Scopus	yes	No APC	-	yes
<a href="#">Scientific Data</a>	Nature	Data Descriptor	WoS PubMed	yes	2490 USD	<a href="#">yes</a>	<a href="#">Blind peer review</a>

Lists of data journals exist online but are not always up to date. See examples of either data journals covering all disciplines [here](#) or specific fields, like [biodiversity](#)



## Data Papers' Pros & Cons

- ✓ Increase the **visibility** of the data you deposit in a repository
- ✓ **Data gets peer reviewed**
- ✓ Scholarly output, fully **citable and indexed** in academic databases  
→ get credit for your work
- ✓ Describe your **data in a human-readable form**, to ensure it's well understood and reused appropriately
- ✓ Increase the FAIRness of your data
- ✗ APCs can be required
- ✗ Takes time to write, submit and publish the data paper



# Data Statements

## Data Statements

*A few sentences included in the published research article providing information about the availability of the data on which the article is based.*



Illustration by Storyset

Some journals have started offering/requiring a dedicated section in the article: “**data availability statement**” or “**data access statement**”. Usually on top or at the end of the article.

Aim: → indicating **where** and **how** to access the underlying datasets, or **explaining why** they can’t be shared freely.

+ **link to their unique identifier !**

# Data statements' examples

**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](https://gitlab.unige.ch/KLab/mis_markers_misplacement).

## 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**.

(Fonseca *et al.* 2020)

(Mouterde *et al.* 2022)

Many publishers provide guidance regarding formulation in their author submission guidelines. See also those examples by the University of Strathclyde Glasgow, covering most cases:

<https://www.strath.ac.uk/research/researchdatamanagementsharing/dataaccessstatements/>

(Lavallard *et al.* 2023)

## Data availability

### 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).

## Data statements' Pros & Cons

- ✓ Increase the **visibility** of the data you deposit in a repository
- ✓ Shows the efforts you have made and your **compliance** attempts with data sharing requirements
- ✓ **Links** your paper with the uploaded datasets in a “**standardised manner**” (machine-extractable but also easy for humans to read)
- ✓ Allows you to provide more information than a simple data citation (e.g. conditions to access, data uploaded in multiples repositories, etc.)
- ✓ No additional cost for you
- ✗ Implies knowing the data link/DOI at the time when you submit your article.



but a lot of repositories now allow to “reserve a DOI”

# Best case scenario

Data uploaded  
in a repository  
(e.g. Yareta)



Gets a good description and  
a unique persistent  
identifier / link (DOI,  
accession number, etc.)

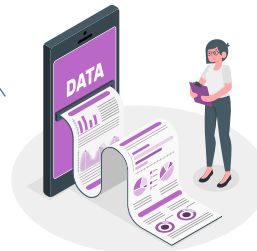
The repository can link back to  
related papers too



A data statement is included in the  
published research paper, pointing  
to the data repository



When uploading your paper in your  
institutional publication repository  
(eg. Archive ouverte UNIGE), indicate  
also the link to the data



A data paper can describe the data  
(if no research paper is published,  
or to complement it)



## Closing remarks

- Uploading data in a repository takes time and work
- There are solutions to make this work visible (findable) and getting credit for it.
- If the uploaded data leads to a paper, include a **data statement** in it!
- If no research publication arises from the data, or if you have time, consider publishing a **data paper**
- Cross-referencing enhances the visibility. Consider **making links everywhere you can.**



# Questions ?



**University of Geneva  
research data team**

[researchdata-info@unige.ch](mailto:researchdata-info@unige.ch)

[Floriane.Muller@unige.ch](mailto:Floriane.Muller@unige.ch)



LOVE  
DATA  
WEEK  
SWITZERLAND



10. – 14. 02. 2025

National Program:

<https://www.researchdatasupport.ch/lovedata>

UNIGE Program:

<https://unige.ch/-/ldw25>





UNIVERSITÉ  
DE GENÈVE

BIBLIOTHÈQUE



Floriane Muller, Bibliothèque de l'UNIGE, 2025

This document is shared under Creative Commons Attribution – ShareAlike 4.0 International : <https://creativecommons.org/licenses/by-sa/4.0>