

# SHARING RESEARCH DATA

Maeve Dwan O'Reilly  
Youth Mental Health Lab

12.05.22



# Agenda

Open Data

Why share data – Wider & personal benefits

When *not* to share data – real & imagined

Metadata

Research project stages

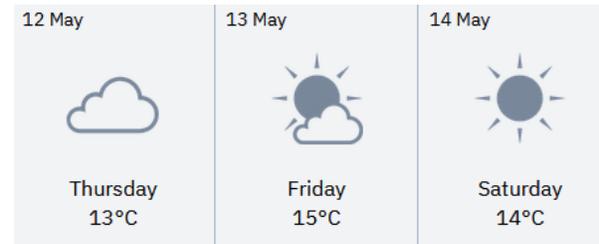
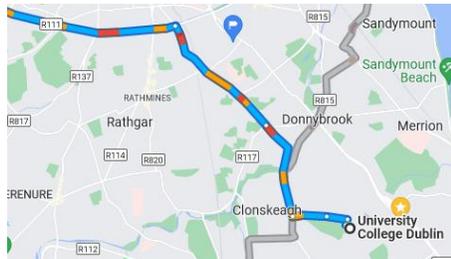
Teacher confidence project



# Open data

*“Open data are data that anyone can access, use and share.” Open Data Institute*

- We live in a world of open data



- While completely open data may not always be possible in our research, we should aim for our data to be **as open as possible, as closed as necessary** (European Commission, 2017).

# Principles guiding data sharing

F

## Findable

Data should be hosted by a stable and recognised open repository.

Data should have a unique persistent identifier.

Data should be accompanied by rich metadata.

A

## Accessible

Data should be published under a licence which facilitates data reuse.

Where data cannot be completely open access, how the data may be accessed should be easy to understand.

I

## Interoperable

Data should be stored in a non-proprietary open file format and described using a standard vocabulary

Users should be able to integrate the data with other data.

R

## Reusable

Data should be understandable

Users should be able to understand the data using only the metadata and documentation provided.

Users should be able to both replicate findings and use the data in studies others than the original.

# Why share data – Wider benefit

Sharing data benefits the public, the research community, participants, and funders:

- Advances science
- Accelerates the pace of discovery
- More efficient research
- Minimises research costs
- Guards against research fraud
- Build the public's trust
- Helps in policy formulation
- Preservation of data in the long term



# Covid-19 and data sharing

## Data sharing done right

- Huge amount of collaboration and data sharing
- How the virus spreads and the speed at which it is spreading
  - [Genome Sequence data bank](#)
  - [Protein Data Bank](#)
- Some datasets shared with articles – 28.5% of PubMed articles

## The harm caused by a lack of data sharing

- Heterogeneity in the way datasets are mentioned, shared, updated and cited
- Papers published but no data or meta data provided
- Little data sharing for some types of data
- Findings shared without data or peer-review
- Only 17.3% of trials declared their intention to share data
- Retraction of papers – falsified data
- Wasteful research?



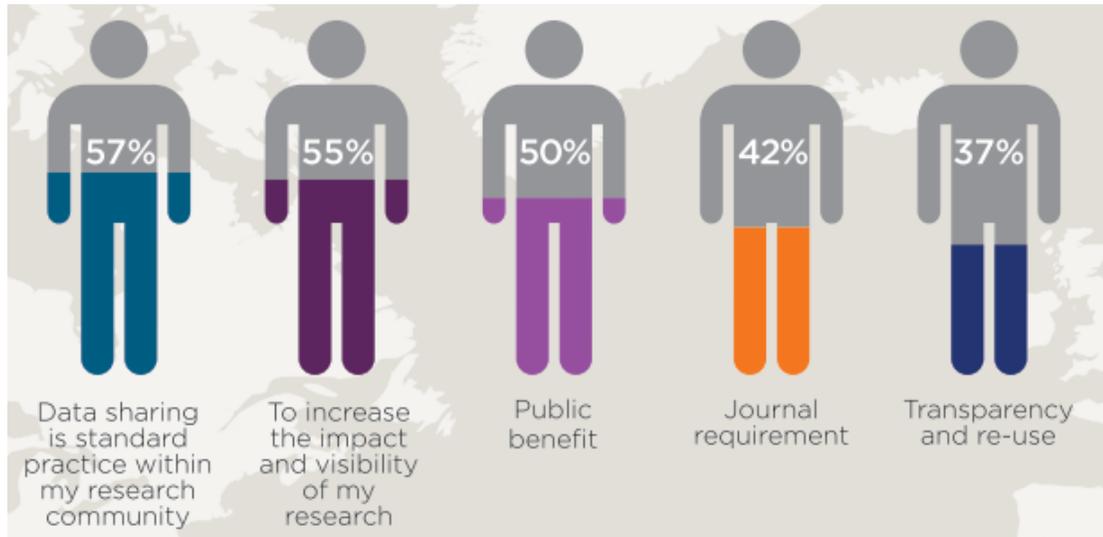
# Why share data – Personal benefits

- Increased visibility of your work
- An archived dataset is a publication of sorts
- Increased citations
- Can lead to collaboration opportunities
- Transparency
- Your data are secure



# Researcher motivations to share research

2014



2019



WILEY

# Why people don't share data (but they should)

Concern	Solution
Privacy concerns and risk to participants	Safeguards in place throughout the research process: Informed consent procedures, good data management, anonymisation and pseudonymisation procedures
Research may be scooped	Don't share your data until after you have published or place an embargo on the use of the data until after you have published.
Misuse or misinterpretation of the data	Provide good documentation to accompany the data. Encourage consultation with future researchers. Deposit the data in a repository that only allows accredited researchers to access the data.
Lack of time and money	Plan data management early in the process. Make data sharing an obligatory part of the process. It is in funders interest that data be shared.
Attribution	Choose where to deposit data carefully. This also requires a change in culture and research policy!

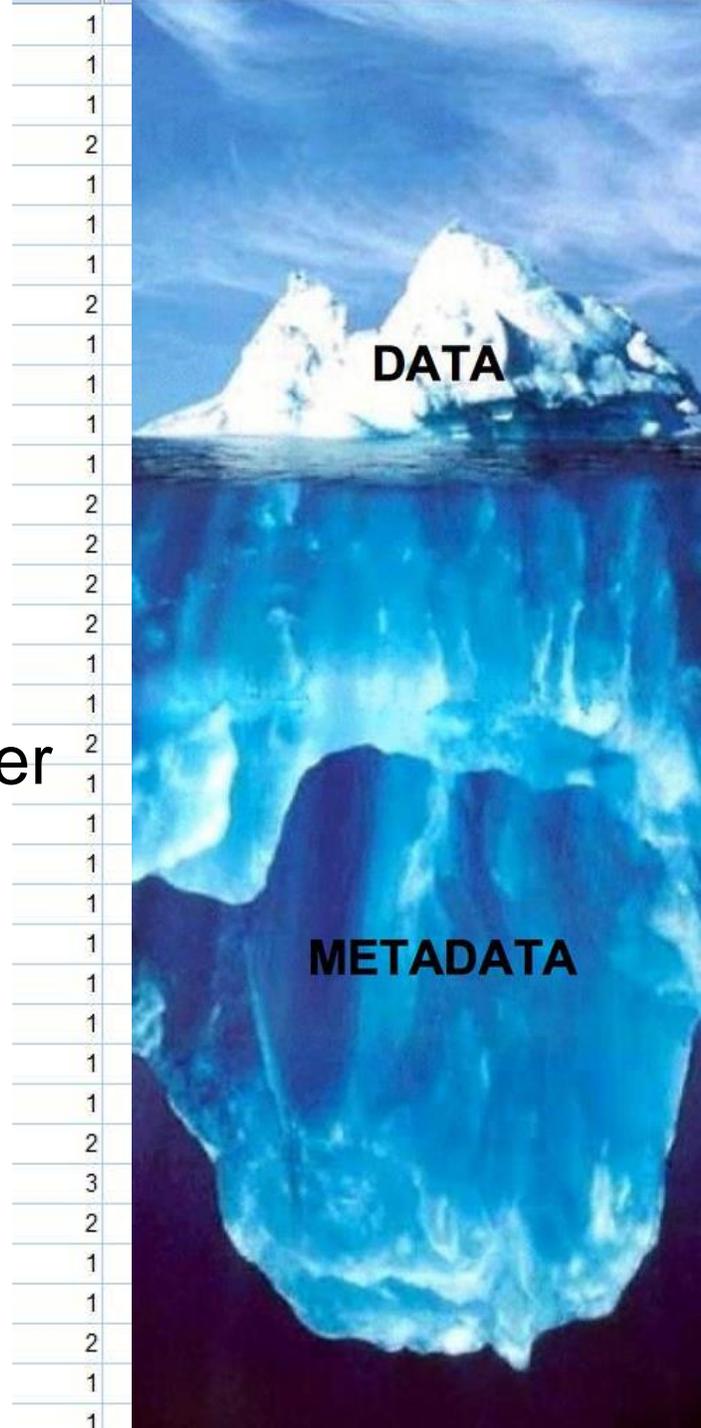


# Metadata

- Data about data
- *“Metadata are a specific subset of data documentation, which provide standardised, structured information explaining the purpose, origin, time references, geographic location, creating author, access conditions and terms of use of a data collection.”* (Corti et al., 2014)
- Metadata are structured to international standards, either generic or discipline specific



- Most data repositories will conform to a standard – you provide the information, they input it.

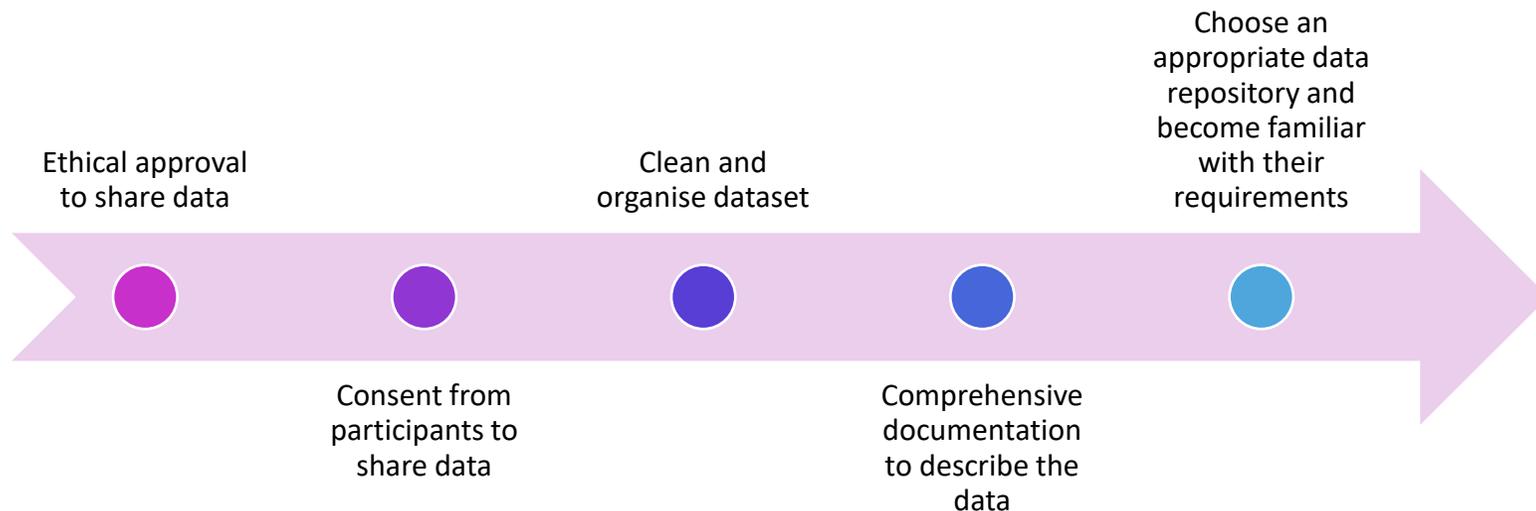


# Research project stages

- Good data management is key at all stages of the research project.

<https://libguides.ucd.ie/data/intro>

- Some things to consider when preparing to share data:



# Repositories

Browse repositories:

- Registry of Research Data Repositories: <https://www.re3data.org/>
- OpenDOAR: <https://v2.sherpa.ac.uk/opensoar/>
- Wellcome Trust approved repositories: <https://wellcomeopenresearch.org/for-authors/data-guidelines#hosting>
- Scientific data (a Nature journal) recommended repositories: <https://www.nature.com/sdata/policies/repositories> (some social science included)

General multidisciplinary repositories:

- DRYAD: <https://datadryad.org/stash>
- Figshare: <https://figshare.com/>
- Zenodo: <https://zenodo.org/>

In Ireland:

- The Irish Social Science Data Archive (ISSDA): <https://www.ucd.ie/issda/>
- The Irish Qualitative Data Archive (IQDA): <https://www.maynoothuniversity.ie/iqda>



# Teacher confidence project

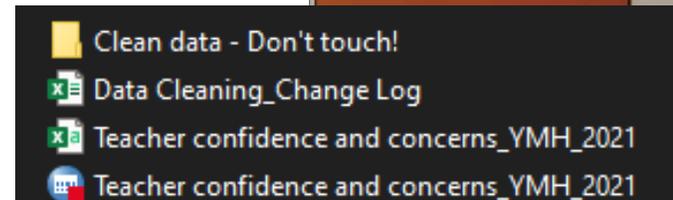
- Cross sectional survey centered around 2 measures; teacher confidence to deliver mental health content and teacher worries about addressing mental health in class.
- N=644 – Secondary school staff in Ireland
- Demographic questions, 6 measures, 4 open questions = 81 variables in the dataset
- Aim to deposit the data with Irish Social Science Data Archive



# Teacher confidence project

- Indicated that the data would be archived once analysis was complete in ethics application form.
- Participant consent form included details of data archiving.

“The anonymous data from this project will be stored securely during the study and will be archived in digital form once the project is complete. This means that other accredited researchers may access it in the future to conduct further analysis or to compare the findings from another project.”
- Kept a record of changes to the measures.
- Cleaned the data set to be as clear as possible.
- Kept a record of data cleaning, including recoding etc.
- Created a master dataset.
- Created a data dictionary according to ISSDA suggestions.
- Completed the data depositor form for the ISSDA.
- Prepared pdfs of the survey and consent form.
- Contacted Andrew Clinch, Data Manager at [issda@ucd.ie](mailto:issda@ucd.ie) with queries



# Preparing a data dictionary

Variable number	Variable name	Variable label	Wording used in survey	Code and meaning
23	SchoolSize	How many students attend your school?	How many students attend your school? <ul style="list-style-type: none"> <li>· &lt;300 students</li> <li>· 300 - 500 students</li> <li>· 501 - 700 students</li> <li>· 701+ students</li> <li>· Does not apply to me</li> </ul>	1=<300 students 2=300 - 500 students 3=501 - 700 students 4=701+ students 5=Does not apply to me
30	TCS_MH_5	I can improve students' general knowledge about mental health.	I can improve students' general knowledge about mental health. <ul style="list-style-type: none"> <li>· 1=Not at all confident</li> <li>· 2</li> <li>· 3</li> <li>· 4</li> <li>· 5</li> <li>· 6</li> <li>· 7</li> <li>· 8</li> <li>· 9</li> <li>· 10=Very confident</li> </ul>	1=1=Not at all confident 2=2 3=3 4=4 5=5 6=6 7=7 8=8 9=9 10=10=Very confident
60	RIBS_Extra_1	In the future, I would be willing to teach a student with a mental health problem.	In the future, I would be willing to teach a student with a mental health problem. <ul style="list-style-type: none"> <li>· Agree strongly</li> <li>· Agree slightly</li> <li>· Neither agree nor disagree</li> <li>· Disagree slightly</li> <li>· Disagree strongly</li> <li>· Don't know</li> <li>· Not applicable</li> </ul>	1=Agree strongly 2=Agree slightly 3=Neither agree nor disagree 4=Disagree slightly 5=Disagree strongly 6=Don't know – recoded as 3 7=Not applicable – recoded as missing

# Preparing a data dictionary

17	Youthreach <sup>1</sup>	Do you work in Youthreach?	Do you work in Youthreach? <ul style="list-style-type: none"><li>· Yes</li><li>· No</li><li>· Does not apply to me</li></ul>	1=Yes 2=No 3=Does not apply to me
----	-------------------------	----------------------------	--	---

<sup>1</sup> “Youthreach” refers to the Youthreach programme/ Youthreach centres run by Education and Training Boards Ireland (ETBI). The Youthreach programme provides education, training, and work experience to early school leavers without qualifications or vocational training that are aged between 15-20 years. For further information: <https://www.gov.ie/en/service/5666e9-youthreach/>

<sup>2</sup> “DEIS” [pronounced: /dɛɪ/] refers to Delivering Equality of Opportunity in Schools (DEIS), An Action Plan for Educational Inclusion which was launched by the Department of Education and Skills in 2005 and later reviewed in 2017. This plan is designed to offer support to schools with high levels of disadvantage. A “DEIS school” in the present survey refers to a school that has been identified by the Department of Education and Skills as needing extra support under the DEIS action plan. For further information: <https://www.gov.ie/en/policy-information/4018ea-deis-delivering-equality-of-opportunity-in-schools/>

## Anonymisation procedure

Variable number	Variable name	How data was anonymised
81	<u>OpenQuestion</u>	Qual data examined, n=1 identifiable response was removed: Email address removed, replaced with “[identifiable data removed]”

# Details of measures

Measure name	Teacher confidence scale for delivering mental health content (TCS-MH)
Authors	Linden & Stuart, 2019
Variable numbers in data set	26-38
Variable names in data set	TCS_MH_1 → TCS_MH_12
Description of measure	12 item measure of school staff confidence to deliver mental health content to young people. Participants were presented with 12 "I can..." statements and were asked to rate their confidence. Example item: "I can answer students' general questions about mental health."
Modifications to the measure	None
Scoring	10-point Likert scale, 1=Not at all confident, 10=very confident, intervening numbers do not have a label. Minimum score=12 Maximum score=120 Higher scores indicate higher levels of confidence
Recoding carried out	None

Measure name	Reported and intended behaviour scale (RIBS)
Authors	Evans-Lacko et al., 2011
Variable numbers in data set	59-66
Variable names in data set	RIBS_1 → RIBS_4; RIBS_Extra_1; RIBS_Extra_2
Description of measure	The RIBS measures stigma by asking participants about their reported and intended behaviours in relation to people with mental health issues. The original RIBS asks participants about their behaviour in four contexts: (1) living with, (2) working with, (3) living nearby and (4) continuing a relationship with someone with a mental health problem. For this study, two items were added to this measure intended behaviour in a school setting. Participants are presented with a statement beginning "in the future I would be willing..." and are asked to rate to what extent they agree with the statement. Example statement: "In the future, I would be willing to live with someone with a mental health problem."
Modifications to the measure	<p>The original RIBS includes four items of reported behaviour, and 4 of intended behaviour. The first four items are not intended to be included in the final score. The study did not include the first four items, only the 4 items of intended behaviour were used.</p> <p>The original scale includes a clarification of "people with mental health problems" as "people seen by healthcare staff" in the scale instructions. This clarification was removed in this study.</p> <p>Two items were added to this measure to identify potential stigma in the school setting: RIBS_Extra_1: In the future, I would be willing to teach a student with a mental health problem. RIBS_Extra_2: In the future, I would be willing to provide support to a student with a mental health problem.</p> <p>A further response option was added for these extra items: Not applicable</p>
Scoring	6-point Likert scale for the 4 RIBS items: 1=Agree strongly; 2=Agree slightly; 3=Neither agree nor disagree; 4=disagree slightly; 5=Disagree strongly 6=Don't know 7-point Likert scale for the 2 RIBS_Extra items: 7=Not applicable Minimum score for original 4-item measure=4 Maximum score for original 4-item measure=20 Minimum score for expanded 6-item measure=6 Maximum score for expanded 6-item measure=30 Higher scores indicate higher levels of stigma.
Recoding carried out	6=don't know recoded as 3=Neither agree nor disagree 7=not applicable recoded as missing

# Data depositor form



## DATA DEPOSIT FORM

### Contents

1. Responsible Parties
2. Collection Description
3. Data and Documentation: Files' Description
4. Details of Methodology and Sampling Procedures
5. Data Transfer

Before completing this form please contact ISSDA to discuss any queries you may have. Please ensure you have consulted the guidance outlined on the **Data Deposit** section of the ISSDA website, in particular the Collection Development Policy and File Format Policy: [www.ucd.ie/issda/deposit](http://www.ucd.ie/issda/deposit)

Metadata

#### Temporal coverage

Dates of fieldwork

Format is MM/YYYY

From: 02/2021

To: 03/2021

Other:

#### Time dimension

Cross-sectional one-time study

Follow-up to cross-sectional study (e.g. longitudinal); please specify number of follow-ups:

Repeated cross-sectional study; please specify how many and how often:

Panel study

Cohort study

Time series

Other, please specify:

#### Geographical Coverage

Country: Ireland

Region: all

County: all

Town: all

Administrative unit:

Other geography:

#### Methods of data collection

Select from list or provide free text using the "other" box:

<input type="checkbox"/>	PAPI (Paper and Pencil Interviewing): With interviewer (face to face, telephone or other)
<input type="checkbox"/>	PAPI (Paper and Pencil Interviewing): Self-completion (distributed by post, email or other)
<input type="checkbox"/>	CAPI (Computer Aided Personal Interviewing)
<input type="checkbox"/>	CATI (Computer Aided Telephone Interviewing)
<input checked="" type="checkbox"/>	CASI (Computer Assisted Self Interviewing). Survey Monkey etc.
<input type="checkbox"/>	Compilation or synthesis of existing material
<input type="checkbox"/>	Other, please specify:



# Conclusion

- We should aim to have our data as open as possible, as closed as necessary and in line with FAIR principles.
- Sharing data has multiple benefits for you personally, the research community, and the wider public.
- Planning to share data begins at the planning stage of a research project.
- BUT it's not a hard or intimidating process.
- UCD Library has excellent resources: <https://libguides.ucd.ie/data/intro>



# References

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library and Information Science Research*, 41(2), 109–122. <https://doi.org/10.1016/j.lisr.2019.04.004>

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. Sage Publication Ltd.

Eurographics. (2019). Putting data sharing platform at heart of national Covid-19 response. <https://eurographics.org/news/ordnance-survey-ireland-puts-data-sharing-platform-at-heart-of-national-response-to-covid-19/>

European Commission Directorate-General for Research & Innovation. (2017). H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. In European Commission (pp. 1–10). [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

Ferguson, L. (2014, November 1). How and why researchers share data (and why they don't). <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-researchers-share-data-and-why-they-dont>

Graf, C. (2019, November 4). What Do Researchers Think About Open Data? <https://www.wiley.com/network/researchers/open-research/what-do-researchers-think-about-open-data>

Graf, C., Flanagan, D., Wylie, L., & Silver, D. (2020). The open data challenge: An analysis of 124,000 data availability statements and an ironic lesson about data management plans. *Data Intelligence*, 2(4), 554–568. [https://doi.org/10.1162/dint\\_a\\_00061](https://doi.org/10.1162/dint_a_00061)

Le Guillou, I. (2020, March 23). Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research. <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/covid-19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research>

Meadows, A. (2014, November 11). To Share or not to Share? That is the (Research Data) Question... . <https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/>

Meyer, M. N. (2018). Practical Tips for Ethical Data Sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131–144. <https://doi.org/10.1177/2515245917747656>

# References

Moir, V. (2019, November 7). What Are the Challenges With Open Research Data? A Primer on FAIR Data .

<https://www.wiley.com/network/researchers/open-research/what-are-the-challenges-with-open-research-data-a-primer-on-fair-data>

NIAID Now. (2021, October 26). Reflections on a Year of COVID-19 Data Sharing . <https://www.niaid.nih.gov/news-events/year-covid-19-data-sharing>

Open Data Institute. (n.d.). What is open data? Retrieved May 9, 2022, from <https://data.gov.ie/edpelearning/en/module1/#/id/co-01>

Piwovar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. PLoS ONE, 2(3).

<https://doi.org/10.1371/journal.pone.0000308>

Tennant, J. (2015, August 26). The Rise of Open Research Data. <https://www.wiley.com/network/researchers/open-research/the-rise-of-open-research-data>

Transparency during global health emergencies. (2020). The Lancet Digital Health, 2(9), e441. [https://doi.org/10.1016/S2589-7500\(20\)30198-9](https://doi.org/10.1016/S2589-7500(20)30198-9)

UCD Library. (2021). Research Data Management: Why Share Research Data . <https://libguides.ucd.ie/data/share>

UGent Data Stewards. (2021, February 20). Knowledge clip: Metadata - YouTube. [https://www.youtube.com/watch?v=DW2T\\_cnqKPU](https://www.youtube.com/watch?v=DW2T_cnqKPU)

Walport, M., & Brest, P. (2011). Sharing research data to improve public health. The Lancet, 377, 537–539. [https://doi.org/10.1016/S0140-6736\(10\)62234-9](https://doi.org/10.1016/S0140-6736(10)62234-9)

Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. Nature Medicine, 28(1), 2–5.

<https://doi.org/10.1038/s41591-021-01654-6>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>

Zuo, X., Chen, Y., Ohno-Machado, L., & Xu, H. (2021). How do we share data in COVID-19 research? A systematic review of COVID-19 datasets in PubMed Central Articles. Briefings in Bioinformatics, 22(2), 800–811. <https://doi.org/10.1093/bib/bbaa331>

# THANK YOU

Maeve Dwan O'Reilly

[maeve.dwanoreilly@ucdconnect.ie](mailto:maeve.dwanoreilly@ucdconnect.ie)

@MaeveDOR 



IRISH  
RESEARCH  
COUNCIL  
An Chomhairle um  
Thaighde in Éirinn

**JIGSAW**  
Young people's  
health in mind

**YMHlab**  
YOUTH MENTAL HEALTH