"I ain't afraid of no myth" – busting the myths on data sharing

Separating the facts from fiction – it is time to silence the myths and shine a light on the truth about data sharing.

In this piece, we cover all angles on the forms and types that data exists as; show you that help is available and who you can turn to for information and guidance; and how data sharing actually establishes and confirms ownership of your data via authorship.

We've broken down the myths and truths into handy sections to help you more easily find the information that you're looking for. We hope this brings you closer to realising the diversity, richness and quantity of data this approach provides.

MedEdPublish

F1000

Me, my data and I – setting the record straight on the wide range of data types and formats to be aware of

"I don't have any data!"
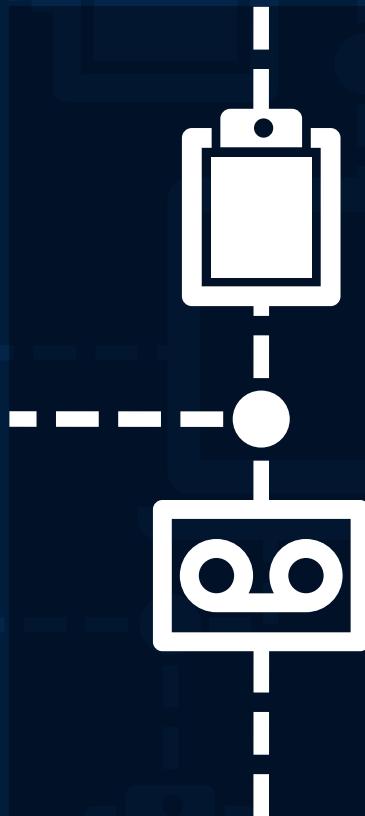
Are you sure? Research data exist in many different forms: **textual, numerical, databases, geospatial, images, audio-visual recordings, data generated by machines or instruments, etc.**

Research data may also include **non-digital materials or 'sources'**. The good news is that some non-digital data can be digitised – for instance, paper can be scanned and analogue audio recordings can be transcribed.

"I don't have any data!"

🔍 More detail

Where data cannot be easily digitised in a way that maintains its usefulness for others, you can still 'share' the data by creating an extensive **metadata record** describing the object, where it is stored, and how to access it.

Depositing this metadata record openly in a repository will allow others to find it. You can cite the metadata record in any associated articles (and vice-versa) in order to establish linkages between the published work and the dataset.

Where published work genuinely has no data associated with it, it's best to indicate this clearly in a data availability statement. This confirms the absence of data to readers, rather than the absence of data sharing.

MedEdPublish

F1000

# "Data sharing isn't a thing in my field"
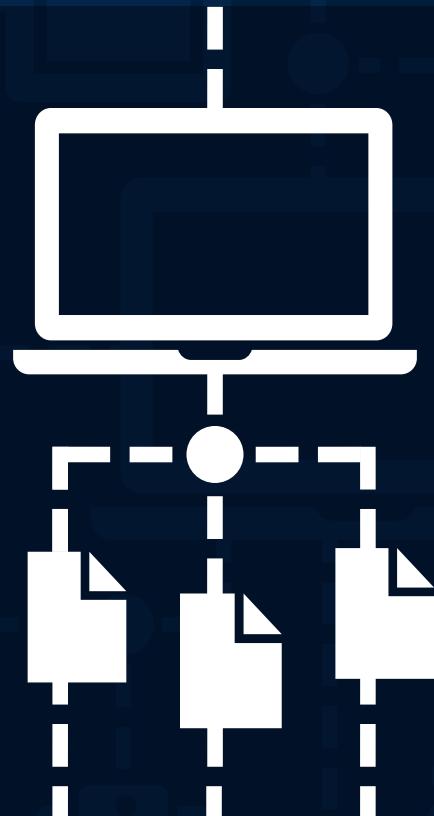
The practice of data sharing, alongside norms and support, varies considerably with each discipline but the key benefits remain the same: reproducibility, credit, and potential reuse.

The **FAIR data principles** can be applied to your research data, regardless of discipline, and there are numerous generalist repositories that accept a wide range of data types, in a wide range of formats. Across disciplines, data sharing mandates are becoming more common at national, funder, and organizational levels, so, this is an opportunity to be a leader in your field.

You are not alone – help is at hand to guide you through the process of sharing your data

"Data sharing is too hard"
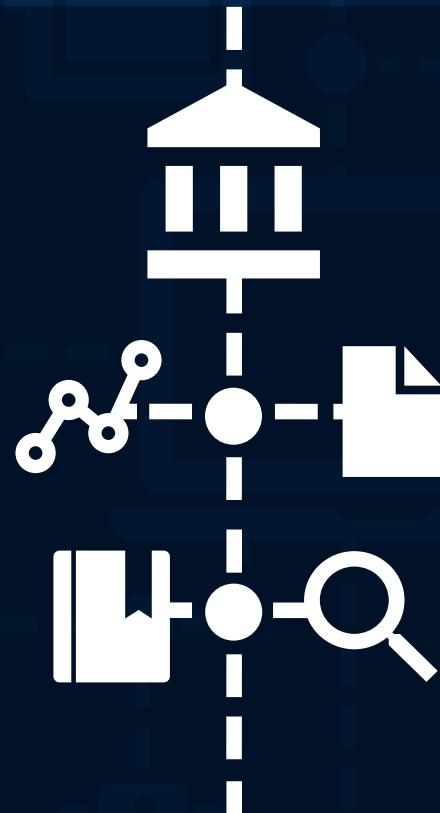
🔍 More detail

Data stewards can help with expert **disciplinary guidance** for data sharing and facilitate early planning using a **data management plan**. Numerous funders also offer services to their grantees to promote data sharing. Beyond this, many funders have started allowing data management to be included in **project budgets**. It's important to **allocate resources** for research data management in future proposals.

MedEdPublish

F1000

"I'm not sure I have the rights to share my data"

Talk to your collaborators.

Early planning of how best to manage research data, including data sharing, is best done using a **data management plan** involving all stakeholders. Using a data management plan will help make it clear **who** has the rights to share data, as well as **how and when**.

Setting aside your concerns on data misuse and how sharing your data gives you ownership and credit

"My data is too sensitive to share"

It's important that data is 'as open as possible, as closed as necessary'.

Depending on what the ethical board approving your study said about data sharing and the level of permission granted from participants, it may be possible to share datasets following appropriate anonymization and/or using controlled access.

"My data is too sensitive to share"

🔍 More detail

Still can't share your data?

Then, **share your metadata!** You can openly publish a description of your data (known as a 'metadata record'). This helps others to discover your data and provides essential information about how the data can be accessed and cited.

For example, you could post a "data codebook" or "data dictionary" in a repository that describes the variables used in your dataset. In this document, you can cite the article in which it appears in order to connect the data description to the paper. Similarly, you can cite the metadata record in your article as part of a **data availability statement**, which should also include the conditions under which your data can be accessed.

MedEdPublish

F1000

"My data will be misinterpreted"
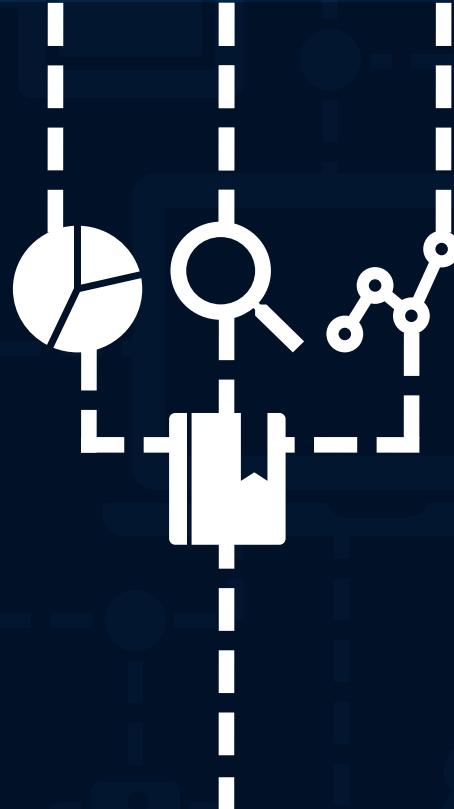
It is important that your data is accompanied by sufficient contextual information to allow others to fully understand your dataset.

A data dictionary is a separate file where each variable is defined, including units and ranges, and often includes other useful information for interpreting the dataset. By helping others (and your future self!) better understand your data, a data dictionary supports reuse and reproducibility – and helps avoid misinterpretation.

"My data will be reused inappropriately"
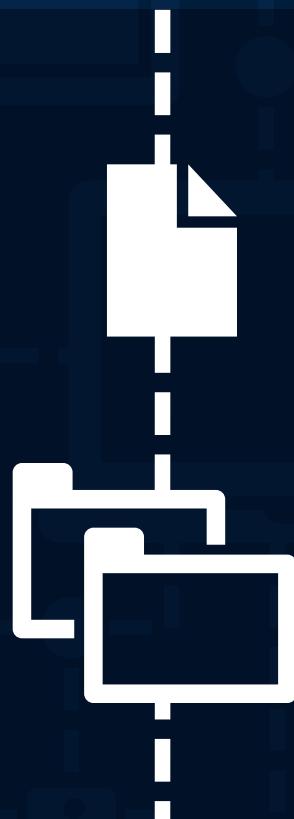
Good **documentation** is key to both stemming and identifying inappropriate use. It is important that your data is supported by **rich metadata** that describes both the purpose of the dataset and any restraints.

Where sensitive data is involved, **data use agreements** make clear the terms under which the data can be used.

# "I am concerned my research will be scooped if I share my data"

The *fear* of scooping is very real, but there is no evidence to support this claim. Instead, data sharing establishes and confirms ownership of your data via authorship.
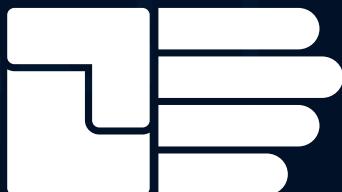
Rather than worry about scooping, we'd encourage you to embrace the possibility of **credit**: where your data is reused by another party, you'll receive recognition through a **formal data citation** (another reason to be sure your data has a DOI!). You may also find new collaborations as a result.

MedEdPublish

F1000

"My data isn't useful to anyone else"
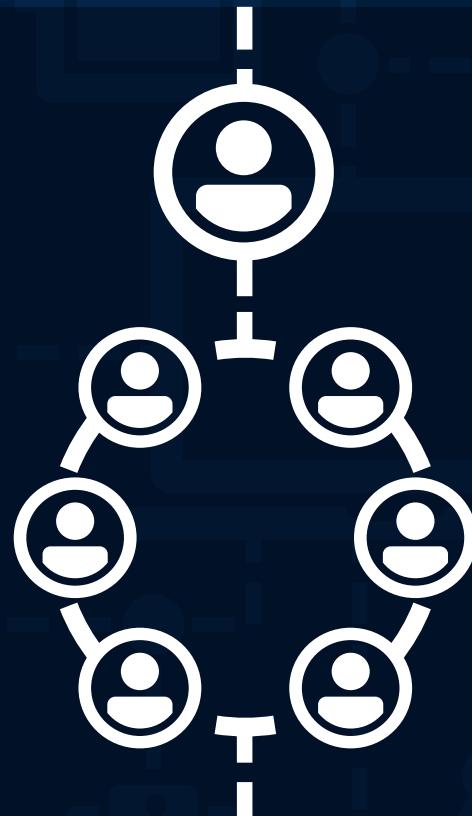
Your research data may have greater reach than you think.

Research data is consumed by a **variety of stakeholders**, beyond researchers, including policymakers and educators. Sharing your data may also help reduce **duplication** of work while promoting **integrative analyses**.

MedEdPublish

F1000

"Sharing my data now will impact my ability to publish later"

The majority of journals and publishers agree that the novelty and value of a research article is not undermined by data sharing.

This means that sharing your data early-on should have **no impact** on your ability to publish a research article later. What's more, there is evidence to suggest that publications associated with a shared dataset have a **citation advantage**.

"Sure, I can share my data. I'll include it as a supplementary file, post it on my webpage, or have others request my data via email"

For data sharing to reach full potential, it's important that data be shared as **FAIR**ly as possible – such that it is most useful to research. This means moving away from traditional modes of sharing where data is difficult to locate and easily lost, and towards **long term preservation** in **trusted data repositories**.

MedEdPublish

F1000

"It does not hurt my research career if I do not share my data"
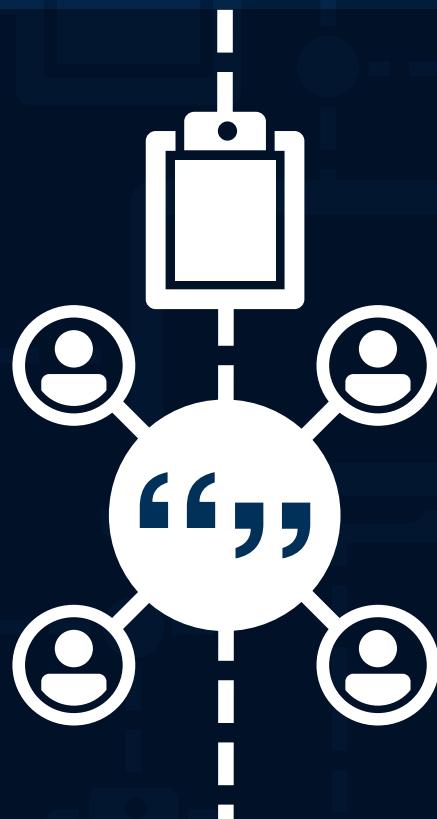
Maybe not – but data sharing has the potential to introduce new collaborations and **increase your citations**. And as the recent public health emergency has highlighted, not sharing your data can have very serious implications.

Alongside a **group of public and charitable** funders, Gates believes that data sharing contributes to: faster progress in improving health, better value for money, and higher quality science.

"It does not hurt my research career if I do not share my data"

🔍 More detail

Plus, with the drive towards Open Science incentive structures are changing. Sharing your data now may be valuable to your career in time.

MedEdPublish