

Open Science and Open Data

Introduction, motivations and how to do it!



Cassandra Gould van Praag (she/her)
Open Science Community Engagement Coordinator

Outline

- 01** The problems which open science (and open data) aim to address
- 02** Solutions (practical and policy) to improve transparency
- 03** Open by design: Preparing your project for data sharing

01

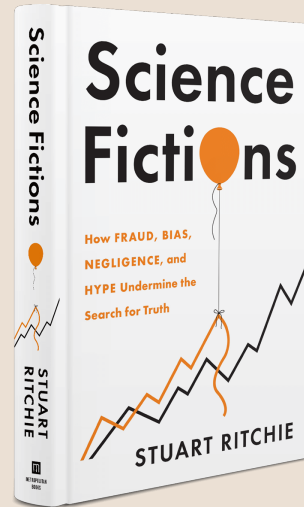
The problems which open science (and open data) aim to address



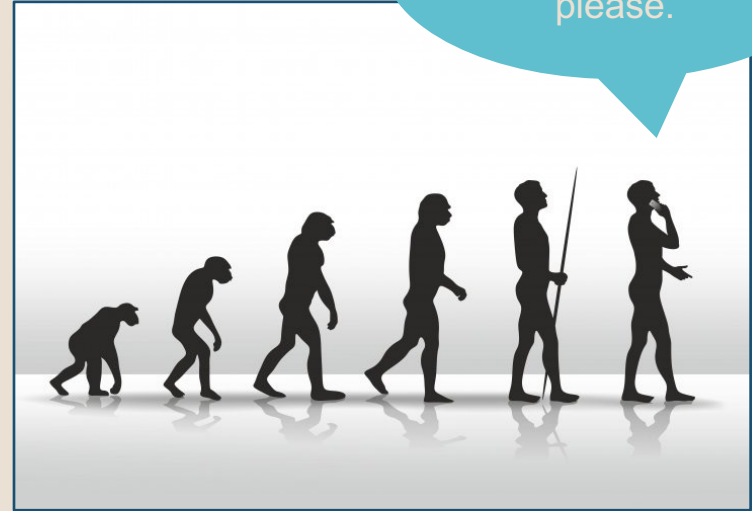
Research is done by humans

We experience and respond to:

- Incentives
- Motivations
- Bias
- Negligence
- Fraud

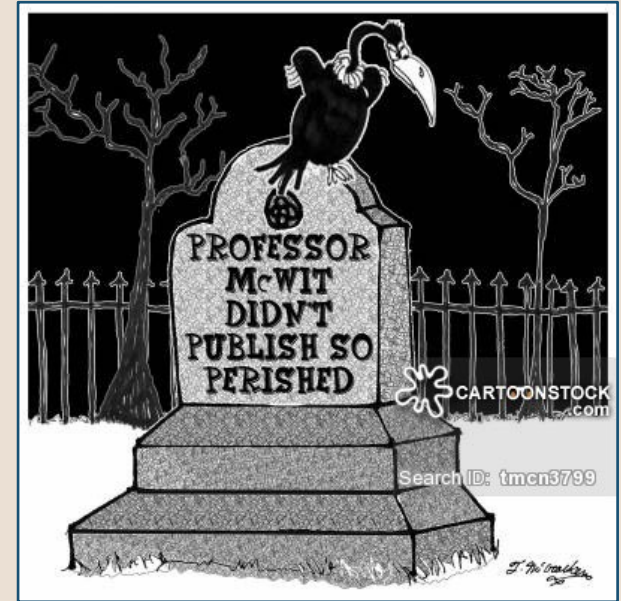


Hello, I'd like to
do a science
please.



Incentive Structures

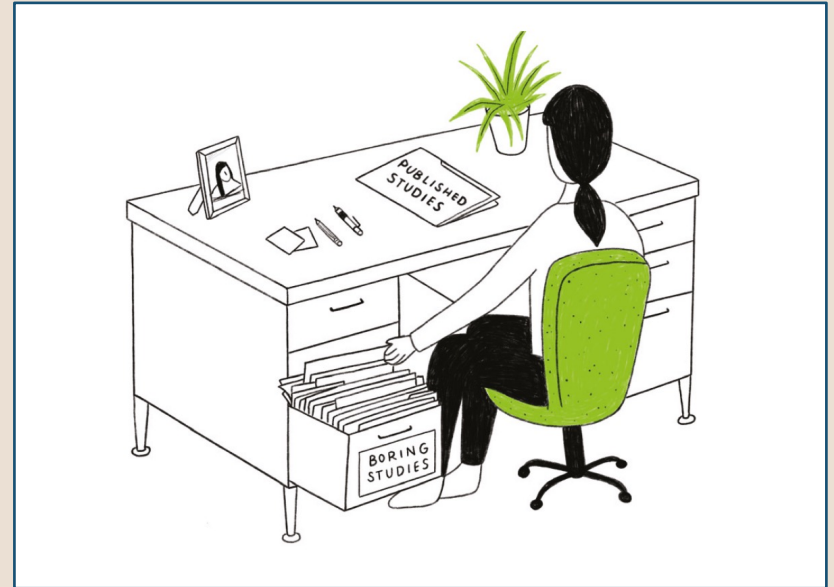
- Publish regularly
- Publish “exciting” results in a high impact journal
- Make sure your work is cited
- Replications are not exciting
- Null results are not exciting
- Tell a story which supports your hypothesis



Publication bias

Randomized clinical trials with positive findings are nearly four times more likely to be submitted and published than trials with negative or null findings.

They are also published 2-3 years earlier than trials with negative findings and more likely to be published in English or in journals with higher impact factor.



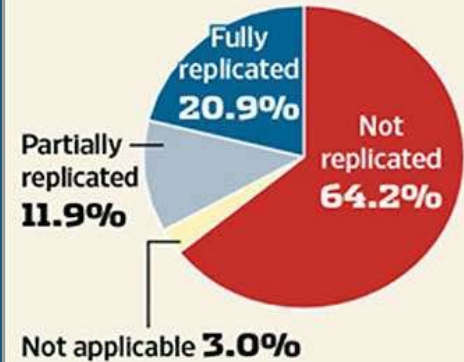
“Reproducibility crisis”

“The success rates for new development projects in Phase II trials have fallen from 28% to 18% in recent years, with insufficient efficacy being the most frequent reason for failure.”

“...validation projects [...] have often resulted in disillusionment when key data could not be reproduced.”

No Cure

When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds failed.

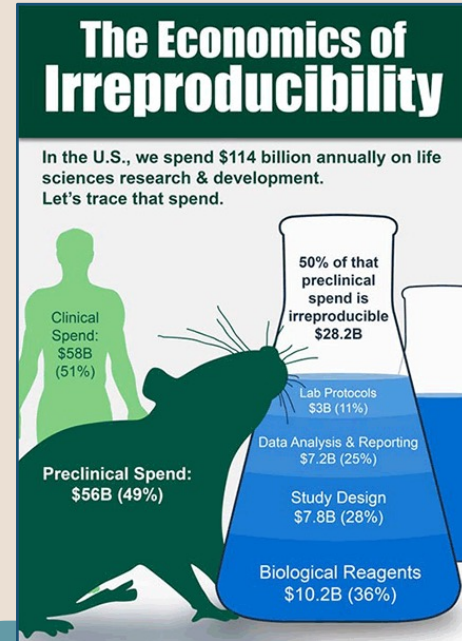


Source: Nature Reviews Drug Discovery

Prinz, F., et al. (2011). *Nat Rev Drug Discovery*

“Reproducibility crisis”

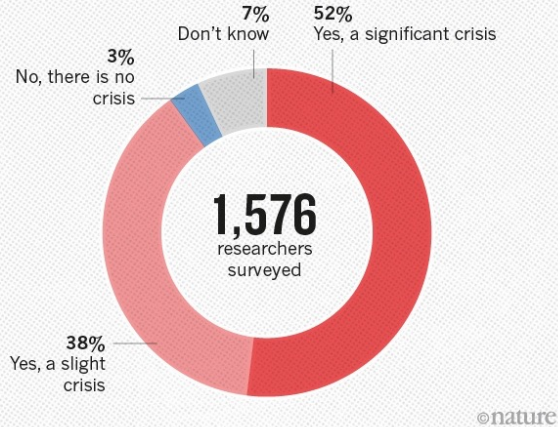
“By now, most scientists are familiar with the global and enormously expensive problem of irreproducible biomedical research.”



“Irreproducibility: A \$28B Year Problem with some Tangible Solutions” (2015). *Genetic Engineering and Biotechnology News*

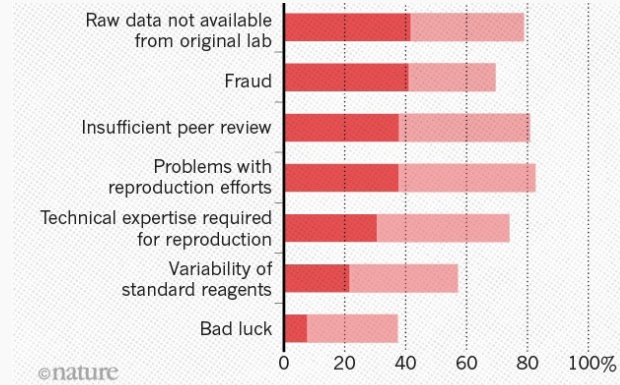
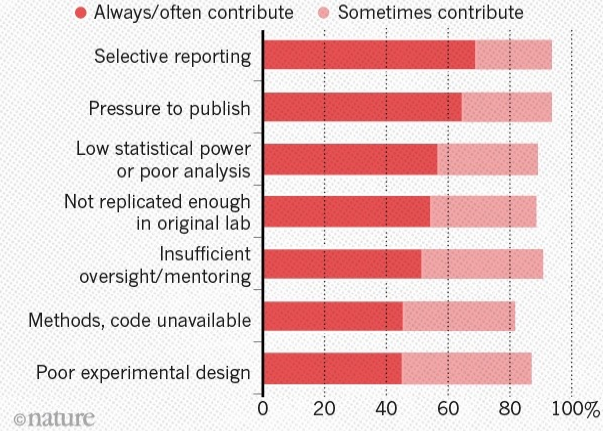
“Reproducibility crisis”

IS THERE A REPRODUCIBILITY CRISIS?



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

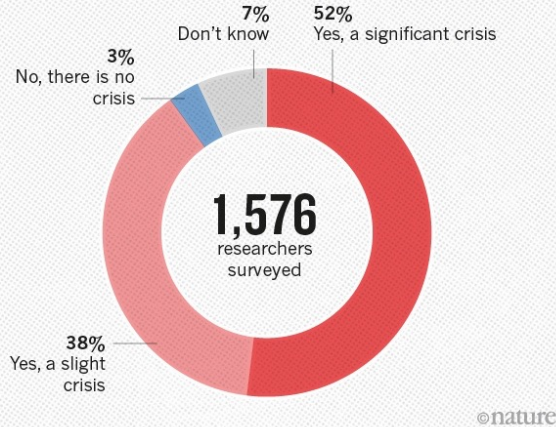
Many top-rated factors relate to intense competition and time pressure.



Baker (2016) *Nature News*

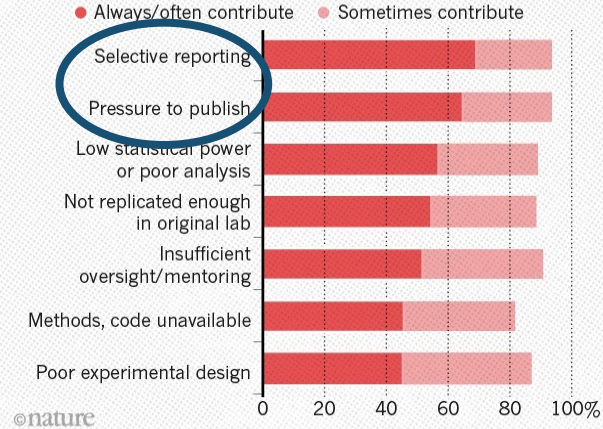
“Reproducibility crisis”

IS THERE A REPRODUCIBILITY CRISIS?

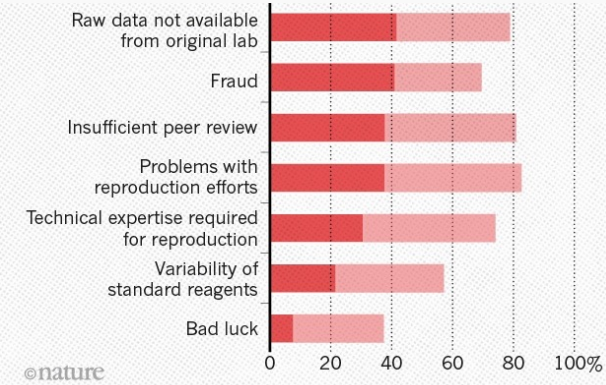


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



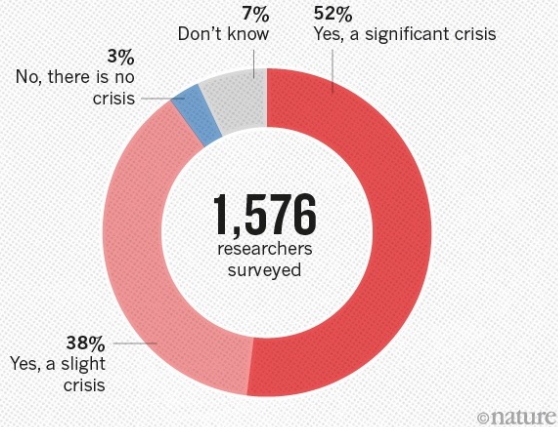
INCENTIVES



Baker (2016) *Nature News*

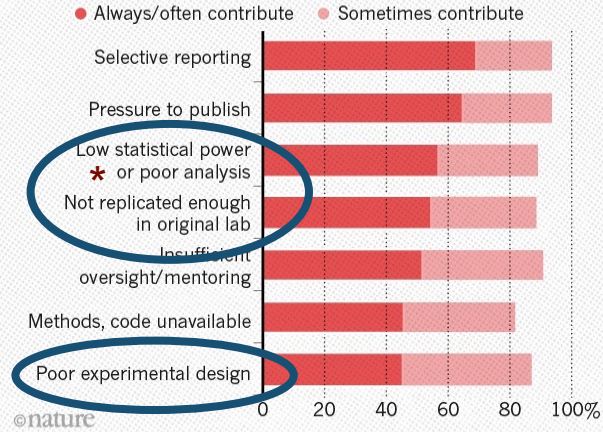
“Reproducibility crisis”

IS THERE A REPRODUCIBILITY CRISIS?

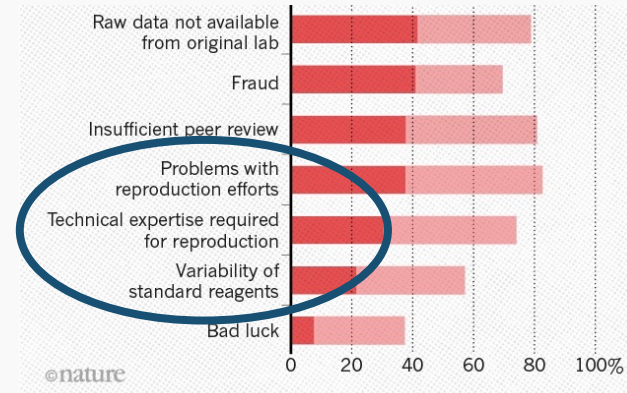


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



METHODS

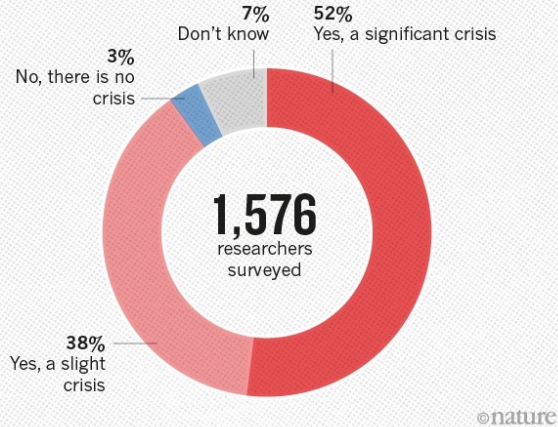


Baker (2016) *Nature News*

*Including analytic flexibility

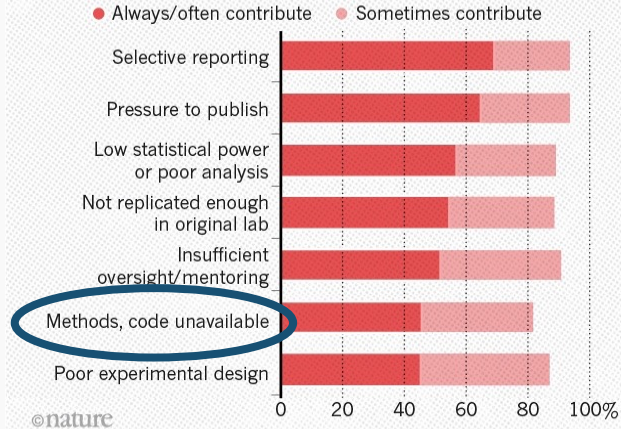
“Reproducibility crisis”

IS THERE A REPRODUCIBILITY CRISIS?

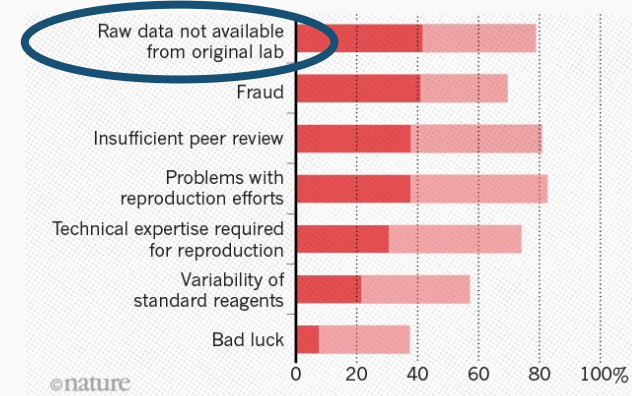


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



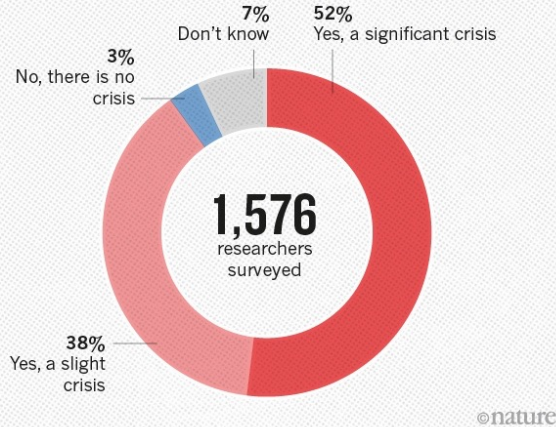
TRANSPARENCY



Baker (2016) *Nature News*

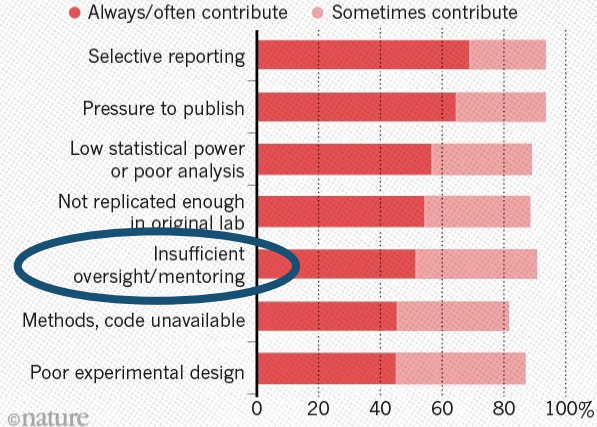
“Reproducibility crisis”

IS THERE A REPRODUCIBILITY CRISIS?

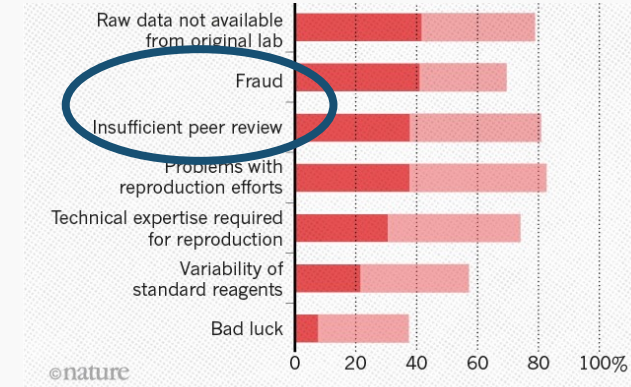


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



CULTURE



Baker (2016) *Nature News*

02

Solutions (practical and policy) to improve transparency



Practical: Open Science

“System of practices for ensuring transparency and reproducibility of research findings.”

(Cass, 2020)



Practical: Open Science

1. Publish (share) all your research outputs
2. Make them Findable, Accessible, Interoperable and Reusable (FAIR)



Policy: Assessment reform



“Eliminate the use of journal-based metrics (e.g. Journal Impact Factors, H-index), in funding, appointment, and promotion considerations.”

“Assess research on its own merits rather than on the basis of the journal in which the research is published.”

Policy: Assessment reform

20,973 individuals and organizations in 153 countries have signed DORA to date.



Oxford University signs Declaration on Research Assessment (DORA)

August 10, 2018 by JulietR

0 Comment

The University of Oxford has signed the [San Francisco Declaration on Research Assessment \(DORA\)](#).

In May 2018, the University's Research and Innovation Committee agreed and accepted the San Francisco Declaration on Research Assessment (DORA) as part of a movement towards responsible use of research metrics at Oxford.



Policy: Funder mandates



“We were the first research funder to introduce a mandatory open access policy.”

“Our policy and guidance on data, software and materials management and sharing sets out what we expect of our researchers.”

“We encourage research outputs to be shared in line with the FAIR principles (findable, accessible, interoperable and reusable).”

Policy: Funder mandates



“UKRI is [...] taking a leading and coordinating role in overseeing the development of open research data policies, infrastructure and services”

“UKRI aims to achieve open research data that is ‘findable’, accessible, interoperable and re-useable (FAIR)”

Policy: Funder mandates

NIHR | National Institute for
Health and Care Research

“The NIHR strongly supports the sharing of data in the most appropriate way.”

Policy: Publisher mandates



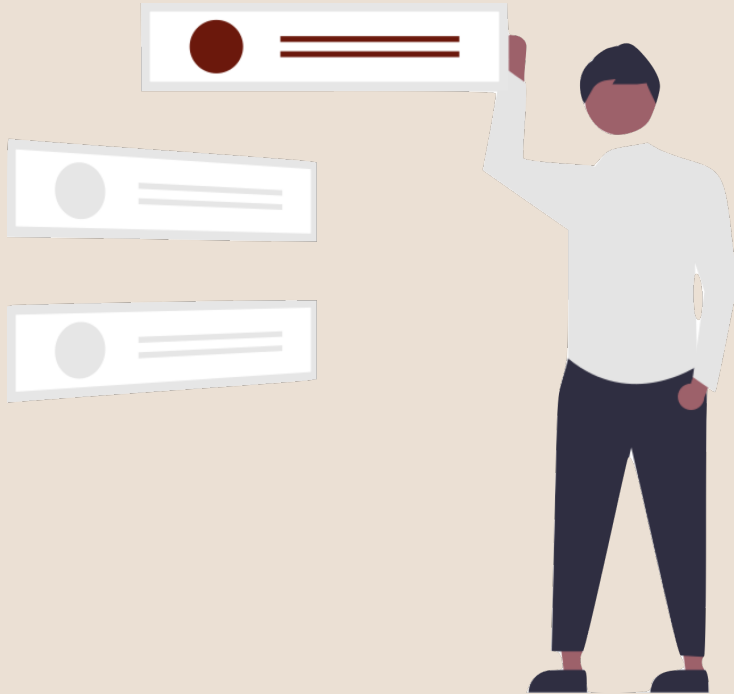
“Research data should be made available free of charge to all researchers wherever possible and with minimal reuse restrictions.”



Enable researchers to achieve “community best practices in the sharing and archiving of research data. We also aim to facilitate compliance with research funder and institution requirements to share data.”

03

Open by design: Preparing your project for data sharing



Open By Design



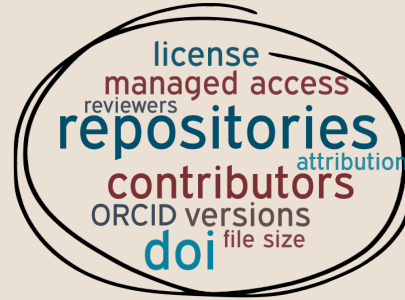
Incorporate sharing into your planning

- What is easy for you to achieve now?
- What is your stretch goal?
- What complications are specific to your work?

Decision framework



Can I share?



How do I share?



Why should I share?



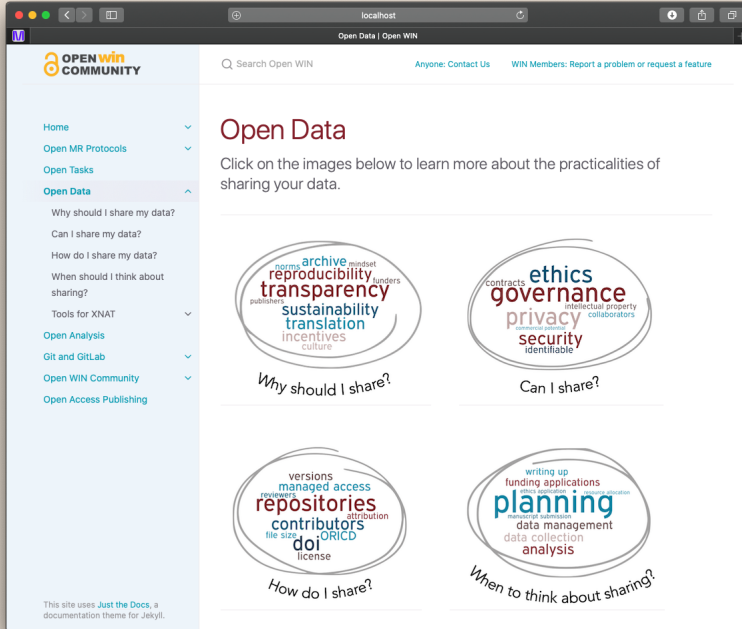
When to think about sharing?



open science



Decision framework



OPEN WIN COMMUNITY

Search Open WIN

Anyone: Contact Us WIN Members: Report a problem or request a feature

Open Data

Click on the images below to learn more about the practicalities of sharing your data.

Why should I share?

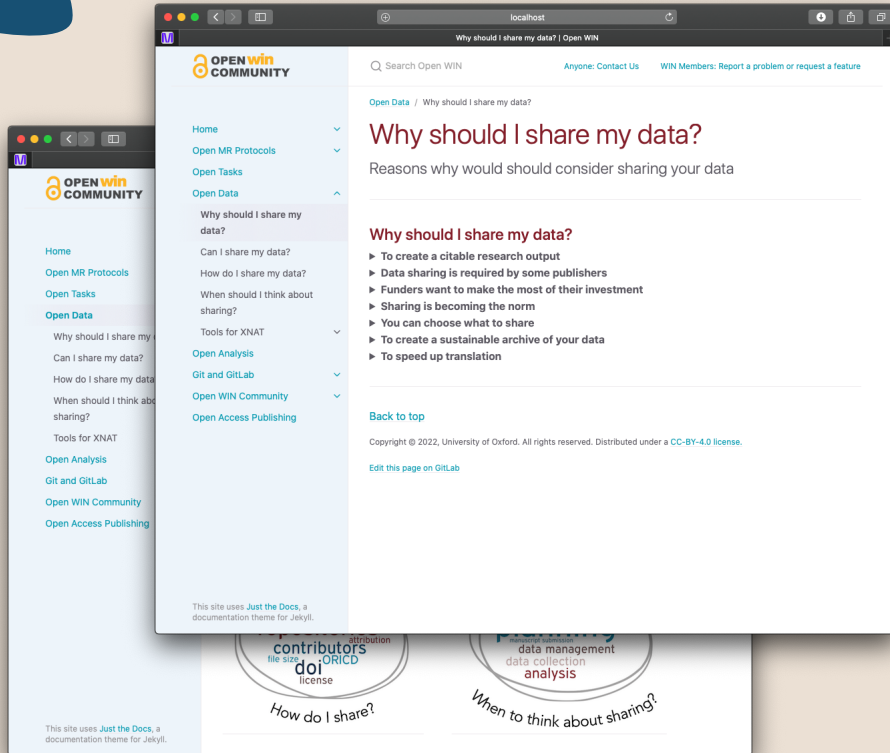
Can I share?

How do I share?

When to think about sharing?

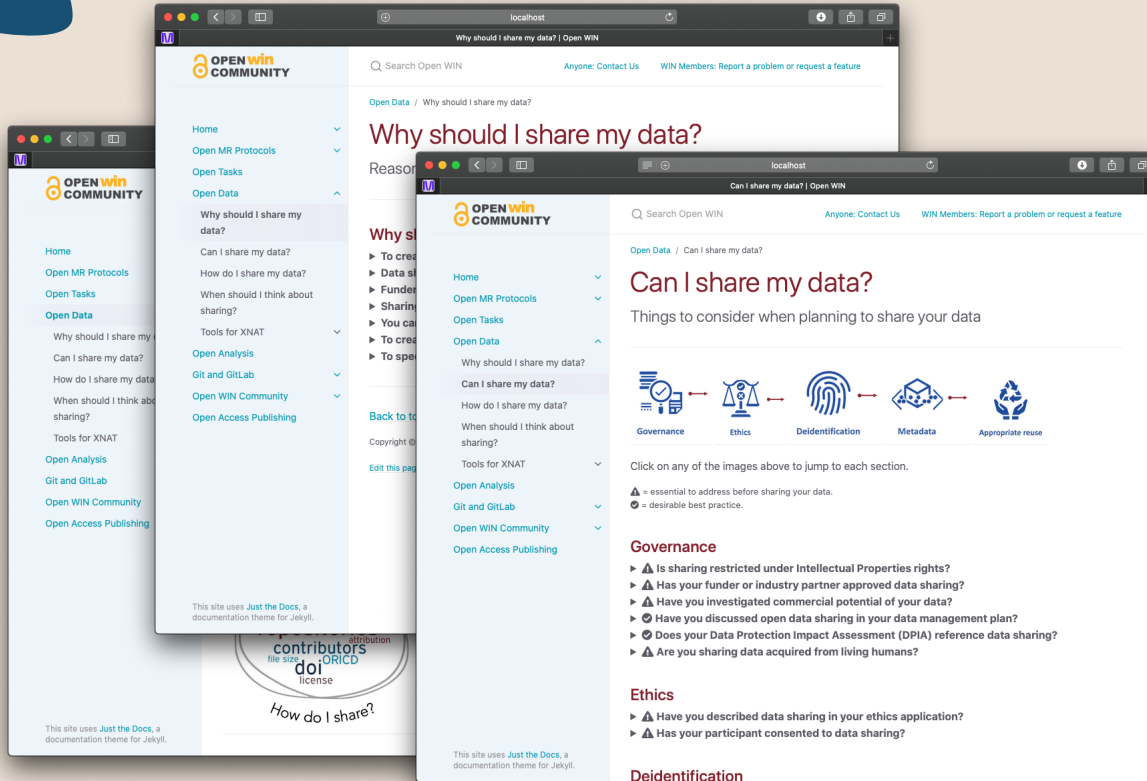
This site uses Just the Docs, a documentation theme for Jekyll.

Decision framework



The image shows a screenshot of a web browser displaying the Open WIN website. The browser address bar shows 'localhost' and the page title is 'Why should I share my data? | Open WIN'. The website has a light blue header with the 'OPEN WIN COMMUNITY' logo and a search bar. A navigation menu on the left lists various topics, with 'Open Data' selected. The main content area features the title 'Why should I share my data?' and a sub-heading 'Reasons why would should consider sharing your data'. Below this, there is a list of reasons: 'To create a citable research output', 'Data sharing is required by some publishers', 'Funders want to make the most of their investment', 'Sharing is becoming the norm', 'You can choose what to share', 'To create a sustainable archive of your data', and 'To speed up translation'. A 'Back to top' link and a copyright notice are also visible. At the bottom of the page, there are two circular icons: one for 'contributors' with 'ORCID' and 'doi license' labels, and another for 'data management' with 'data collection' and 'analysis' labels. The text 'How do I share?' and 'When to think about sharing?' is written below these icons.

Decision framework



The image displays three overlapping browser windows showing the Open WIN website. The top window shows the 'Why should I share my data?' page. The middle window shows the 'Can I share my data?' page with a navigation diagram. The bottom window shows the 'Can I share my data?' page with detailed sections for Governance, Ethics, and Deidentification.

Why should I share my data?

Reasons to share your data

- To create new knowledge
- Data sharing can help you to
- Funders and funders want to
- Sharing your data can help you
- You can get credit for your
- To create new knowledge
- To speed up research

Back to top


Copyright © 2022 Open WIN

Edit this page

This site uses [Just the Docs](#), a documentation theme for [Jekyll](#).

Can I share my data?

Things to consider when planning to share your data



Click on any of the images above to jump to each section.

- ⚠ = essential to address before sharing your data.
- 👍 = desirable best practice.

Governance

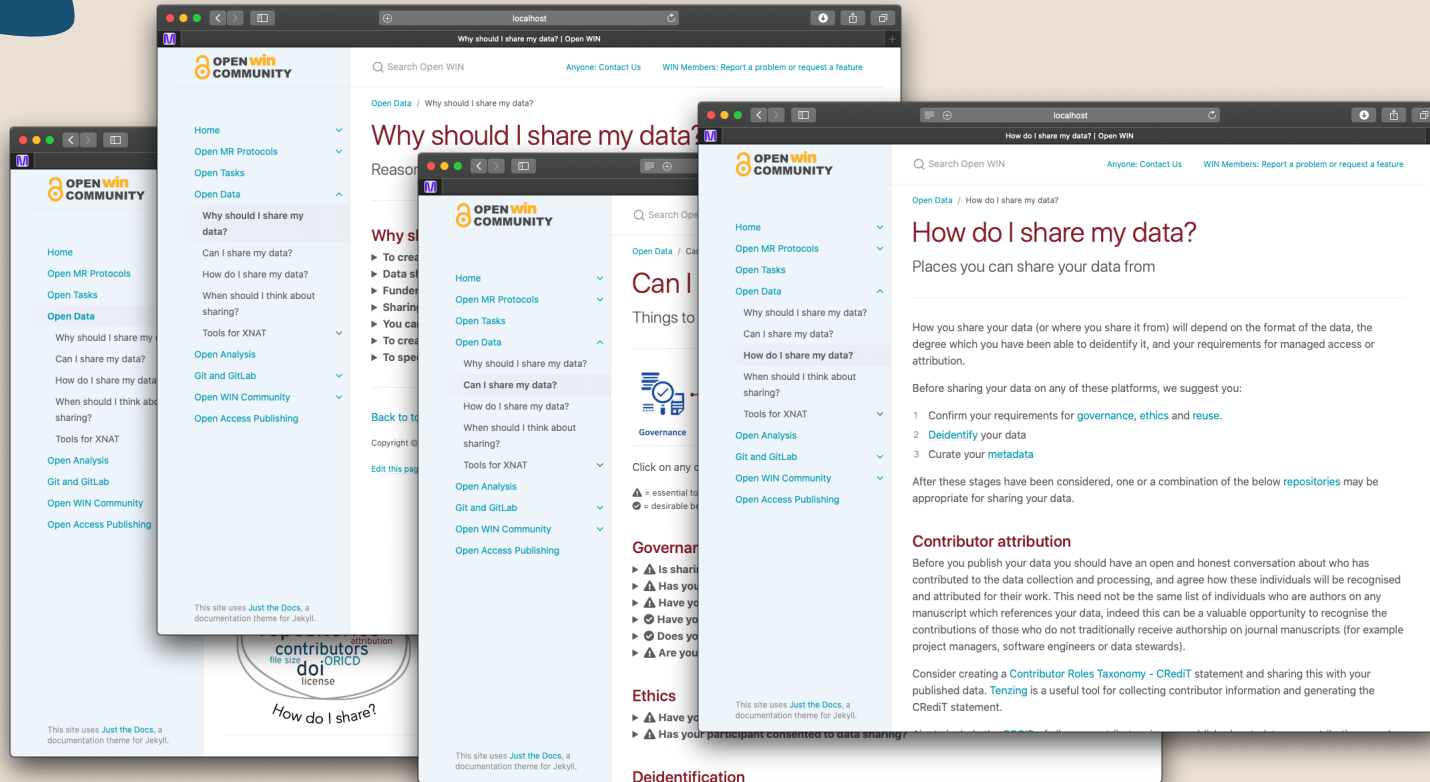
- ⚠ Is sharing restricted under Intellectual Properties rights?
- ⚠ Has your funder or industry partner approved data sharing?
- ⚠ Have you investigated commercial potential of your data?
- 👍 Have you discussed open data sharing in your data management plan?
- 👍 Does your Data Protection Impact Assessment (DPIA) reference data sharing?
- ⚠ Are you sharing data acquired from living humans?

Ethics

- ⚠ Have you described data sharing in your ethics application?
- ⚠ Has your participant consented to data sharing?

Deidentification

Decision framework



OPEN WIN COMMUNITY

Search Open WIN

Anyone: Contact Us WIN Members: Report a problem or request a feature

Open Data / Why should I share my data?

Home

Open MR Protocols

Open Tasks

Open Data

Why should I share my data?

Can I share my data?

How do I share my data?

When should I think about sharing?

Tools for XNAT

Open Analysis

Git and GitLab

Open WIN Community

Open Access Publishing

Why should I share my data?

Reasons

Why should I share my data?

► To create

► Data sharing

► Funding

► Sharing

► You can

► To create

► To create

Back to top

Copyright ©

Edit this page

This site uses [Just the Docs](#), a documentation theme for [Jekyll](#).

How do I share my data?

Places you can share your data from

How you share your data (or where you share it from) will depend on the format of the data, the degree which you have been able to deidentify it, and your requirements for managed access or attribution.

Before sharing your data on any of these platforms, we suggest you:

- 1 Confirm your requirements for [governance](#), [ethics](#) and [reuse](#).
- 2 [Deidentify](#) your data
- 3 Curate your [metadata](#)

After these stages have been considered, one or a combination of the below [repositories](#) may be appropriate for sharing your data.

Contributor attribution

Before you publish your data you should have an open and honest conversation about who has contributed to the data collection and processing, and agree how these individuals will be recognised and attributed for their work. This need not be the same list of individuals who are authors on any manuscript which references your data, indeed this can be a valuable opportunity to recognise the contributions of those who do not traditionally receive authorship on journal manuscripts (for example project managers, software engineers or data stewards).

Consider creating a [Contributor Roles Taxonomy - CRediT](#) statement and sharing this with your published data. [Tenzing](#) is a useful tool for collecting contributor information and generating the CRediT statement.

Can I share my data?

Things to consider

Governance

- **Is sharing**
- **Has your**
- **Have you**
- **Have you**
- **Does your**
- **Are you**

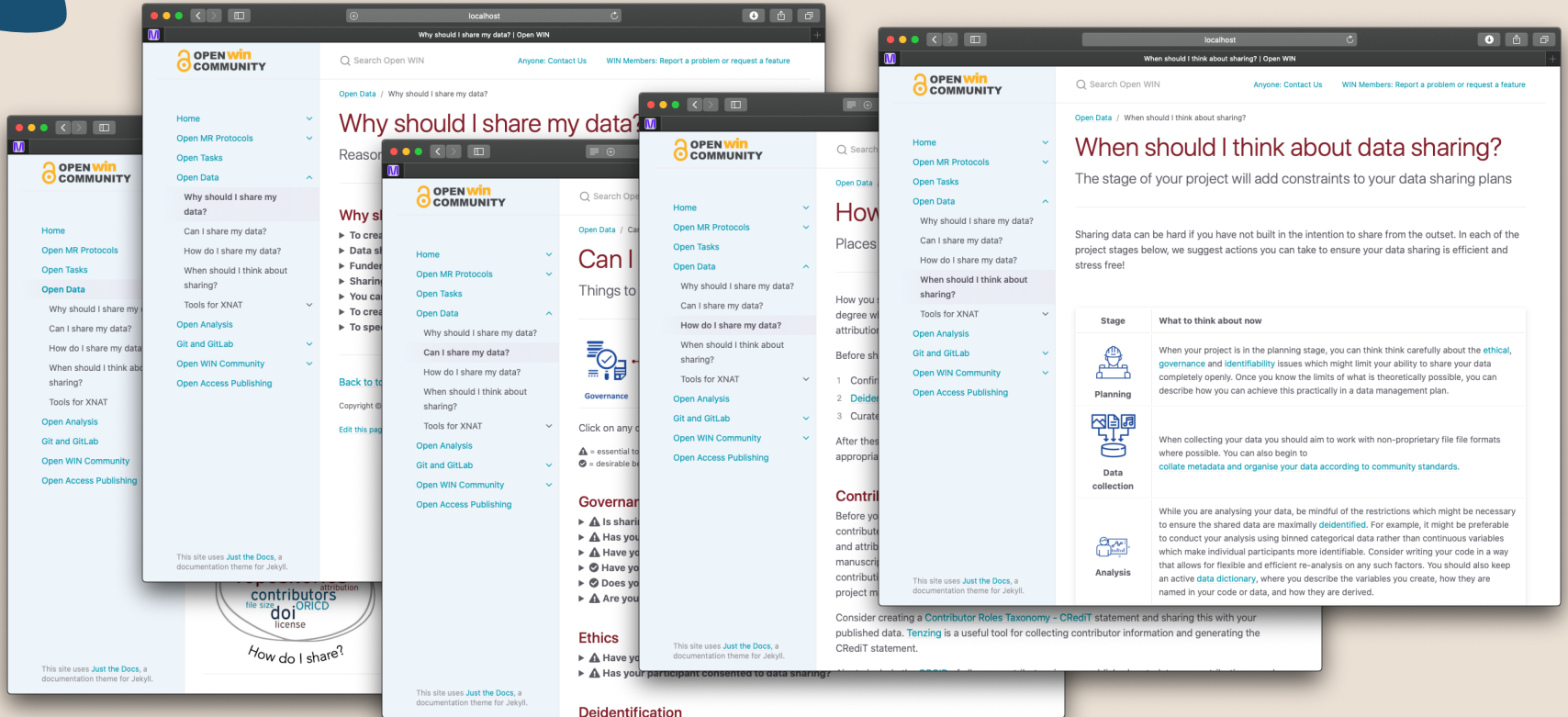
Ethics

- **Have you**
- **Has your participant consented to data sharing?**

Deidentification

This site uses [Just the Docs](#), a documentation theme for [Jekyll](#).

Decision framework



The image shows a collage of overlapping browser windows displaying the 'OPEN win COMMUNITY' website. The windows are arranged to show different sections of the site, illustrating a decision framework for data sharing. The visible content includes:

- Home** and navigation menus (Open MR Protocols, Open Tasks, Open Data, Open Analysis, Git and GitLab, Open WIN Community, Open Access Publishing).
- Why should I share my data?** - A page with a search bar and a list of reasons.
- When should I think about data sharing?** - A page with a search bar and a table of stages and considerations.
- How to share my data?** - A page with a search bar and a list of steps.
- Can I share my data?** - A page with a search bar and a list of questions.
- Ethics** - A page with a search bar and a list of ethical considerations.
- Governance** - A page with a search bar and a list of governance topics.
- Data sharing** - A page with a search bar and a list of data sharing topics.
- Deidentification** - A page with a search bar and a list of deidentification topics.
- Contributor Roles Taxonomy - CRediT statement** - A page with a search bar and a list of contributor roles.

The website uses a clean, modern design with a light blue and white color scheme. The navigation menu is on the left side of each page. The main content area is on the right. The footer of each page includes the text: "This site uses Just the Docs, a documentation theme for Jekyll."

Conclusions

- 01** “Closed” science is susceptible to human biases and error
- 02** Publishing research outputs improves transparency and reproducibility
- 03** Decision framework for designing open projects: Why, Can, How, When

Your questions

1. When do you need a contract to share data?
2. Is it legal to share data through publications?
3. What types of data can I share?
4. What platforms exist for sharing?
5. How much control do we have about what is shared, to whom, and when?

Your questions

1. When do you need a contract to share data?
 - When it isn't anonymous, otherwise it is GDPR Personal Data
 - Anything in participant space is unique to the individual and *potentially* identifiable
 - If anonymous but you want to restrict reuse with specific requirements (e.g. Authorship)

Your questions

1. When do you need a contract to share data?

- CC license options



- **CC BY:** This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, **so long as attribution is given to the creator.** The license allows for commercial use.



- **CC BY-SA:** This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, **you must license the modified material under identical terms.**



CC0 (aka CC Zero) is a public dedication tool, which allows creators to give up their copyright and put their works into the worldwide public domain. CC0 allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, **with no conditions.**



- **CC BY-NC:** This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format **for noncommercial purposes only,** and only so long as attribution is given to the creator.



- **CC BY-ND:** This license allows reusers to copy and distribute the material in any medium or format **in unadapted form only,** and only so long as attribution is given to the creator. The license allows for commercial use.

Your questions

1. When do you need a contract to share data?

- WIN Data Usage Agreements (DUA) will:
 - Explicitly inhibit re-identification
 - Give you authorship control
 - Give you “downstream” sharing control
 - Impose funder specific requirements (e.g. Acknowledgements)

Your questions

2. Is it legal to share data through publications?

- Yes, when:
 1. The data are anonymous, AND/OR
 2. You are confident that you have the right license in place, have arranged a contract, or when using one of WINs Data Usage Agreements

Your questions

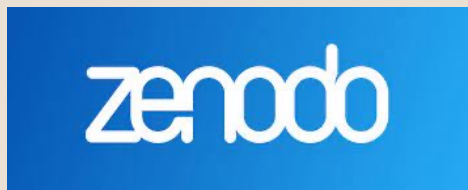
3. What types of data can I share?
 1. Raw (with correct contract / licence)
 2. Maximally de-identified (with correct contract / licence)
 3. Anonymised participant level
 4. Derived participant level
 5. Values used to generate your figures!

Your questions

4. What platforms exist for sharing?



- All data types, anonymous only, any CC license, embargo, reviewer only links, doi



- All data types, anonymous only, any CC license, embargo, reviewer only links, doi + doi versions

Your questions

4. What platforms exist for sharing?



- MR data only, anonymous (statistical maps), CC-0 only
- MR, MEG, EEG, PET, anonymised only (normalised space), CC-0 only, requires BIDS

Your questions

4. What platforms exist for sharing?



- Any data and DUAs can be created. Long process for ethical and privacy review.



- Non-human, post mortem, “conditions of use” specified by the researcher

Your questions

5. How much control do we have about what is shared, to whom, and when?
- Legally, as much as you want if/when you specify the terms.

Your questions

1. When do you need a contract to share data?
2. Is it legal to share data through publications?
3. What types of data can I share?
4. What platforms exist for sharing?
5. How much control do we have about what is shared, to whom, and when?