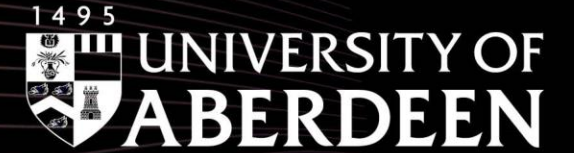


ExplORe Series

Making Data Open: What you need to know



Lesley MacRae
21st February '23



Why should I make my data open?

- Good Science! : Add to the knowledge pool
- Have your data used in other's research
- Meet funder/institutional requirements
- Have your work confirmed, debunked, extended
- Avoids p-hacking and harking

“A key study published in PLOS One in 2020 found a correlation of **up to 25.36% more citations** for articles that share their data in a repository”

How sharing your data could increase your citations, Springer Nature, Jun 08, 2021

Great way to get more exposure, especially as an ECR...

Reproducibility and replicability?

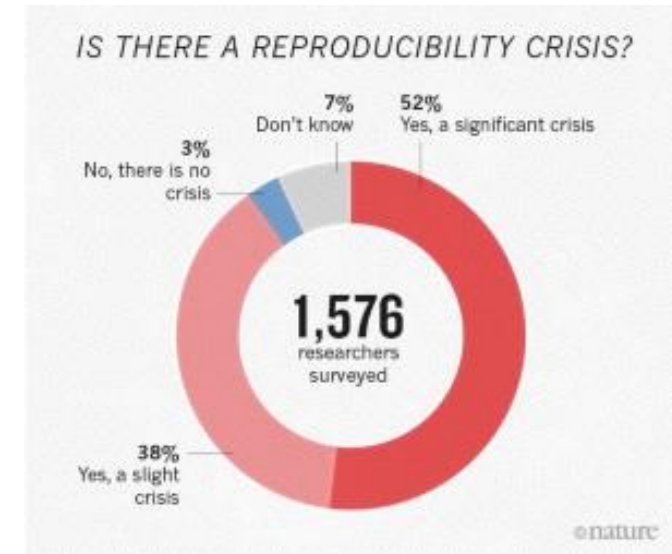
Reproducibility :

The same analysis with same data/code – but performed by a different person

Replicability :

Re-performing the experiment but using your own data – to create your own results

“More than half of researchers have tried and failed to reproduce another scientists experiments, and more than half have failed to reproduce their own experiments” — Baker, Nature 2016



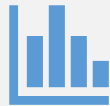
I'm a bit worried about data sharing

- My data might be misinterpreted
- Is my data interesting enough
- People might contact me to ask questions
- My data is in a mess
- I'm too busy....
- Scooping



Next... **What** to share... and **How** to share it

What to deposit?



Data (open/common file formats)



Metadata (data about data)



Any other documentation (codebooks, software code, dataset structure)

How? Firstly...Plan

- Data Management Plan
- Keep track of everything
- Pre-register
- Share and license
- Report transparently, including all data and code in articles

Most researchers say their data sharing practice evolves and gets more organized and rigorous as their career develops

Creating a DMP

- Check course booking Digital Research run online training sessions

[Data Management Plan | University Systems and Software | Toolkit | The University of Aberdeen \(abdn.ac.uk\)](#)

- DMPonline provides funder specific examples
- For more support contact : digitalresearch@abdn.ac.uk



- Public DMPs of EC-funded projects : [DMP Use Case Project \(University of Vienna Phaidra - o:1140797\) \(univie.ac.at\)](#)
- Public DMPs in DMPOnline, argos and zenodo



Why pre-register?

- Prevents HARKing
- Prevents P-hacking (data dredging)
- Publication bias



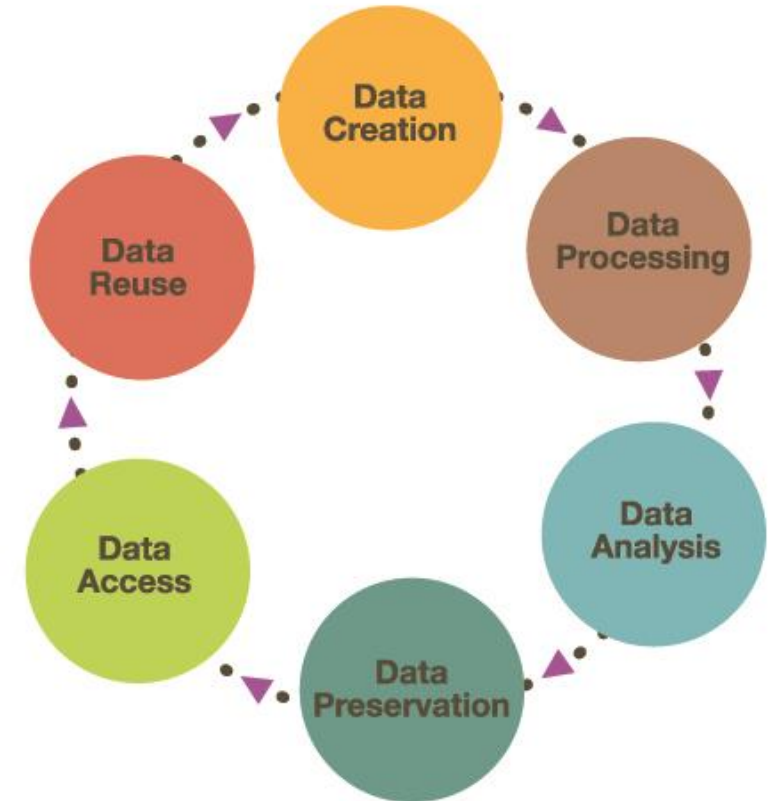
Where to pre-register

- As Predicted
- OSF Pre-registration
- Subject specific platforms



In an ideal world...

1. Data Management Plan, Pre-register
2. Collect and organize your data
3. Consider publishing in a pre-print
4. Publish open access
5. Share your data/methods/code with an open license



<https://libguides.ntu.edu.sg/rdm/researchdatalifecycle>

FAIR

FAIR \neq Open

as open as possible,
as closed as
necessary

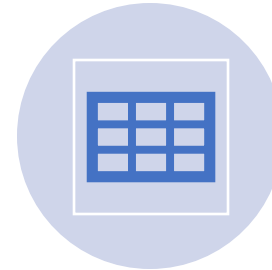


Image: 'Balancing rocks' by Viewminder CC-BY-SA-ND www.flickr.com/photos/light_seeker/7780857224

FAIR principles



To make data digestible for humans and machines



The principles apply to metadata (data about our data) as well as data



Not only intended for open data



Principles NOT standards and can be applied gradually to your work

What FAIR means: 15 principles



Findable:

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the identifier of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;

Interoperable:

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

Accessible:

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol;
 - A1.1** the protocol is open, free, and universally implementable;
 - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary;
- A2.** metadata are accessible, even when the data are no longer available;

Reusable:

- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes;
 - R1.1.** (meta)data are released with a clear and accessible data usage license;
 - R1.2.** (meta)data are associated with detailed provenance;
 - R1.3.** (meta)data meet domain-relevant community standards;

doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Slide CC-BY by Erik Schultes, Leiden UMC

Comprehensive descriptions can be found at <https://www.go-fair.org/fair-principles/>

F A I R

Findable Accessible Interoperable Reusable

- Metadata
- PIDs
- Repositories

- Metadata
- Open file formats and software

- Metadata
- Ontologies
- Repositories

- Metadata
- Licences

Metadata, a minimum...

- Title of the dataset
- Creators (contact details)
- Identifier –DOI
- Funder (grant details)
- Rights (licensing info)
- Access info (embargoes?)
- Language
- Project dates
- Project description
- Subject (keywords)
- Methodology
- Data structure
- Variable names

README.txt file – File folder hierarchy/context for the data

MIT Metadata Guidance :
<http://libraries.mit.edu/data-management/store/documentation/>

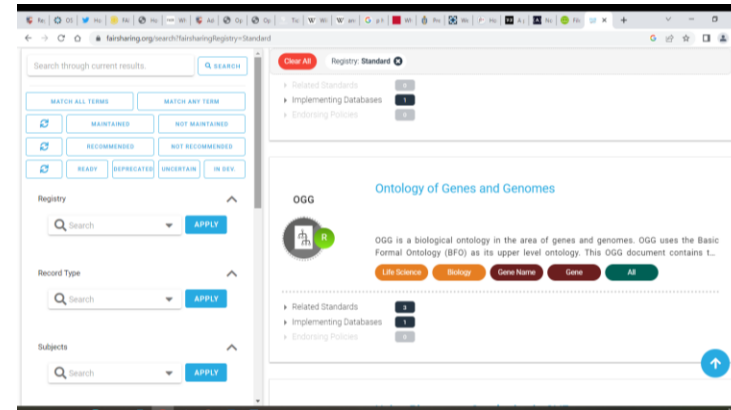
Useful resources for FAIR data

Fairsharing.org

- search for repositories, funder policies and standards
- lots of discipline specific policy information

ARDC FAIR checklist

- <https://ardc.edu.au/resource/fair-data-self-assessment-tool/>
- Check if your data complies with FAIR and tips to make it more compliant



Australian Research Data Commons

Repositories

Publish Data – Where?

Journal service for supplementary material

Meet publisher requirements

Data available from published results

It can be costly and risky with data rights

Closed and unlikely access to ensure preservation

Institutional data repository

Accept various types of data, ensure long-term access

More reliable and there will be no costs

May not offer long-term sustainable access

May not have disciplinary metadata

Generic repository

Reach a wider audience.

Accepts several types, suitable for interdisciplinary data

Usually only simple metadata is available

No editorial control over the quality of deposited materials

Disciplinary repository

Offers expertise and experience in data management

Likely to accept complete data sets

Selective in the type of data they accept

Requires planning and high standards, may incur costs

Journal requirements



Language | English

Journal Article Publishing Support Center

Journal Article Publishing Support Center > Submission > What are the requirements for my supplementary material?

All Topics Search

Research & Preparation

Login & Account

Submission

Review

What are the requirements for my supplementary material?

Last updated on April 11, 2022
Requirements for supplementary material for your supplementary material



Disciplines Products Resources About



- Home
- Publishing your research
- Choose Open
- Research impact
- Policy s
- Event s
- Insights blog
- Researcher Services

Home > How to publish your research > Writing your paper > Enhancing your article with supplementary material

Enhancing your article with supplementary material

Search: keyword, title, author



Additional to the

Journal Authors/Editors/Reviewers

Journal Author Gateway

How to Get Published Resources

Manuscript Submission Guidelines

Ethical Statements Guidance

Supplemental Material - Guidelines for authors

Supplemental Material – Guidelines for authors

The following guidelines outline SAGE's general policy for handling supplementary material, but please always refer to individual journal author guidelines to check for any bespoke policies.

Supplemental material refers to files related to a specific article, which authors supply for publication alongside their article. They should generally be additional pieces to the article that could not be included in the issue or print version, such as appendices, tables, and audio and video material that is impossible to produce within the article. There may also be instances, such as certain podcasts and videos, where they are not directly attributable to a specific piece of research.

For information on the specific handling and hosting of research data files, and SAGE's open research data policies, please click [here](#).

PURE Institutional Repository



- Links to your profile
- There is a size limit if using for storage
- Always submit your data details to PURE

A screenshot of the University of Aberdeen PURE Institutional Repository website. The browser address bar shows 'abdn.ac.uk/staffnet/research/access-pure-14608.php'. The page features the University of Aberdeen logo and navigation menus. A purple 'StaffNet' header contains links for 'Working Here', 'Policy and Governance', 'Teaching and Learning', 'Research and Knowledge Exchange' (highlighted), 'News and Events', and 'Staff Directory'. Below this is a breadcrumb trail: 'University Home / StaffNet / Research and Knowledge Exchange / Pure / Access Pure'. A left-hand menu lists: 'University Home', 'StaffNet', 'Research and Knowledge Exchange', 'Pure' (highlighted), 'Access Pure', 'Research Profiles', 'Adding Content', 'Fix Your Content', and 'FAQs and Glossary'. The main content area is titled 'Access Pure' and includes a login link: 'Log into Pure at <https://pure.abdn.ac.uk/admin/login.xhtml>'. Below this is a section titled 'Information on Accessing Pure' with a paragraph of text. At the bottom, there is a 'Logging in to Pure' section.

Generic repository

- Wider audience
- Interdisciplinary
- May only allow simple metadata



DRYAD



OSF

zenodo



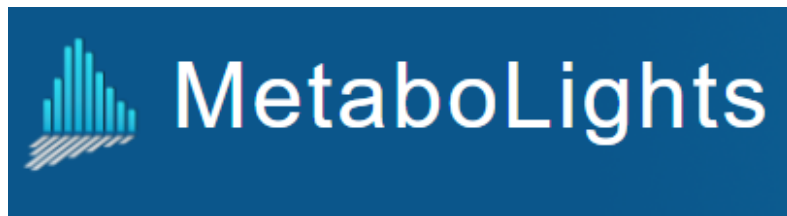
figshare

The
Dataverse[®]
Project

[Generalist-Data-Repository-Grid.pdf \(agu.org\)](#)

Disciplinary repository

- Offer expertise
- Likely to accept complete datasets
- May be more selective in the data they accept



Number of Submissions



Data paper Journals

- Scientific Data (Nature)
<https://www.nature.com/sdata>
- Data in brief (Elsevier)
<https://www.journals.elsevier.com/data-in-brief/>
- Data (MDPI)
<https://www.mdpi.com/journal/data>
- Patterns (bio data intensive science)
<https://www.cell.com/patterns>

SCIENTIFIC DATA



Data in Brief

> Open Access



data

Patterns

Trustworthy Repositories

- Certification Tools for repository auditing & certification.
- CoreTrustSeal (CTS)
- Nestor Seal;
- ISO 16363: 2013



Digitalbevaring.dk

Sharing research data, Pedro Principe, OpenAIRE 7/17/22

Licensing data and code

Open licenses

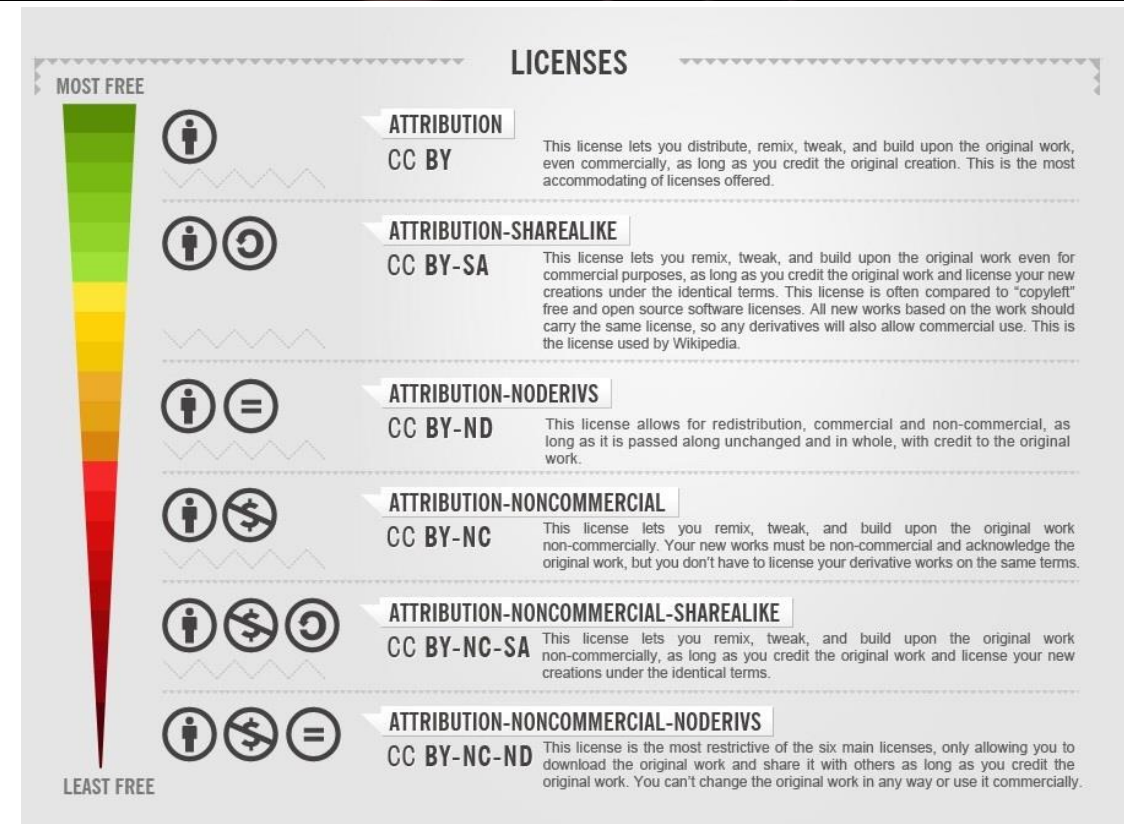
- Applying a license removes uncertainty about what others can do with your work
- You can grant re-uses without others having to contact you
- You can modify the conditions and infringements you want to place
- Be clear! The easier you make it for people the more likely they will respect your wishes

Licensing Datasets

1. Check if the data can be © : Not all data is copyrightable in all jurisdictions. If the data is not copyrightable or if you are not sure, apply a CCO or other open data public domain dedication. This provides clarity for the reuser.
2. Check who owns ©: If you don't own the rights, you need to comply with the requirements of the original rights holder.
3. If you own the © : Determine what license you want to apply.
 - Remember that any license imposing more conditions than attribution makes re-use, crawling, text and data mining either very complicated or simply impossible
 - Consider applying CC0 or other open data public domain dedication. If you want to ensure you are correctly attributed, provide a clear statement in the credits.

Creative Commons

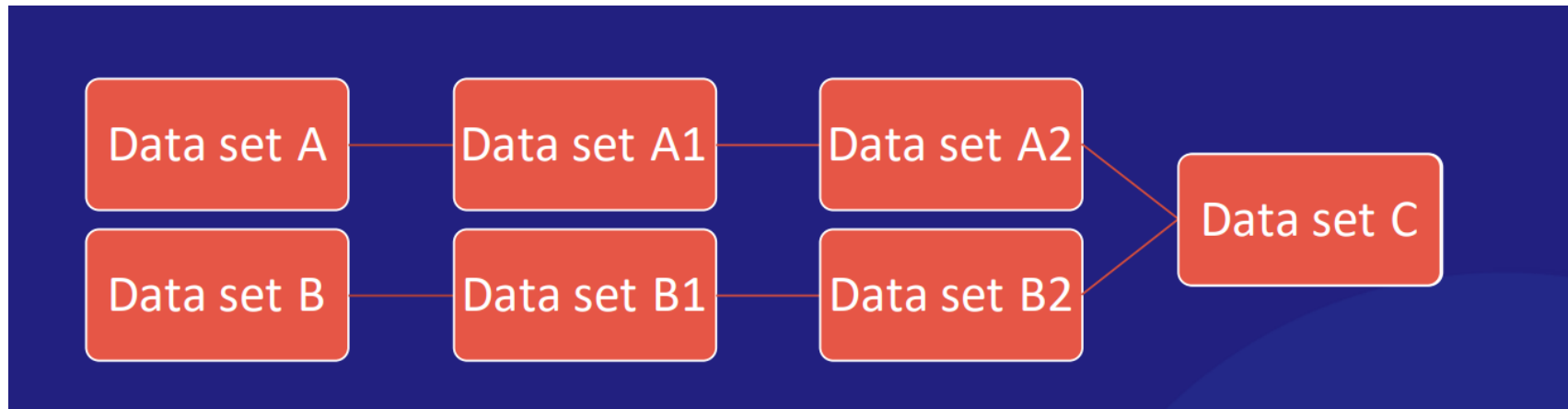
- CC licenses are universally recognisable
- Easily applicable
- Easy to change from the default of all rights reserved to some rights reserved



<https://creativecommons.org/licenses/by-sa/3.0/>Foter Creative Commons license diagram

For data...

- Attribution – involves giving credit and linking to the license. But to cite data set C you would need to cite all the other sources.
- This can lead to attribution stacking!



For data ...

- ShareAlike – states that you must distribute remixes under the same license as the original. But that may be incompatible with other material the user wants to use in their remix
- No Derivatives – if you remix or build upon the material, you can't distribute the modified material. But with a dataset its impossible to extract and use just parts of it

CC0 – open license

Free to :

Share – copy and redistribute in any medium or format

Adapt – Remix, and build on, even commercially

No Attribution required – You're free to give appropriate credit (best practice) but its not required



CC0 is the most appropriate licence for sharing data. It avoids attribution stacking, allows commercial reuse, and allows meta-analyses to be carried out

Creative Commons at a glance

Good for

- very simple, factual datasets
- data to be used automatically

Watch out for

- versions: use v. 4 or later
- attribution stacking
- the NC condition: only use with dual licensing
- the SA condition as it reduces interoperability
- the ND condition as it severely restricts reuse

[Creative Commons Attribution 3.0 Unported](#)

Licensing code

Proprietary – source code kept carefully guarded

Permissive – All rights given, including right to re-license (MIT, Apache)

Copyleft (in between) – others can copy and distribute as long as they track changes in the source code (GPL licenses)

Licensing code

	Public domain & equivalents	Permissive license	Copyleft (protective license)	Noncommercial license	Proprietary license	Trade secret
Description	Grants all rights	Grants use rights, including right to relicense (allows proprietization, license compatibility)	Grants use rights, forbids proprietization	Grants rights for noncommercial use only. May be combined with copyleft.	Traditional use of copyright ; no rights need be granted	No information made public
Software	PD, CC0	BSD, MIT, Apache	GPL, AGPL	JRL, AFPL	proprietary software, no public license	private, internal software
Other creative works	PD, CC0	CC-BY	CC-BY-SA	CC-BY-NC	Copyright, no public license	unpublished

https://en.wikipedia.org/wiki/Permissive_software_license

Choosing a licence

Choose a License

Answer the questions or use the search to find the license you want

Start again

What do you want to deposit?
Software Data

Search for a license...

Public Domain Mark (PD)
The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.
Publicly Available PD

Public Domain Dedication (CC Zero)
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.
Publicly Available CC BY ND OPEN DATA

<https://ufal.github.io/public-license-selector/>

<https://creativecommons.org/choose>

<http://choosealicense.com>

<https://tldrlegal.com>

GNU General Public License v3 (GPL-3)

Code License managed by kevin, submitted 9 years ago. #Copyleft #GNU #OSI-Approved #Open Source

Summary Fulltext Changesets 512143

Quick Summary

You may copy, distribute and modify the software as long as you track changes/dates in source files. Any modifications to or software including (via compiler) GPL-licensed code must also be made available under the GPL along with build & install instructions.

Can	Cannot	Must
<ul style="list-style-type: none">Commercial UseModifyDistributePlace Warranty	<ul style="list-style-type: none">SublicenseHold Liabile	<ul style="list-style-type: none">Include OriginalState ChangesDisclose SourceInclude License

Thank You