

FAIR Data Checklist – Mechanical Engineering¹

The FAIR (Findable, Accessible, Interoperable, Reusable) principles were created to more effectively share data with both people and computers. This checklist and FAQs are designed to help you ensure your data is FAIR. A more complete overview of FAIR found at www.go-fair.org. While this is mainly tailored towards digital datasets, the principles can also be applied to software and physical data.

Permission to share data:

- ☐ All the data is mine to share
- or*
- ☐ I have permission from the owners to share any external data
- ☐ I have linked to/referenced any data already publicly available
- _____
- ☐ I have considered whether I need to embargo any data
- ☐ I have considered any export control² or intellectual property issues

Personal data (where relevant):

- ☐ I have participant consent to share any personal data
- ☐ The data is anonymised / pseudonymised where possible
- ☐ The data is being shared according to the data management plan in my ethics review

Data:

- ☐ My data is in an open or commonly used format where practical
- ☐ My files are logically organised
- ☐ My data is as complete as possible (both raw and processed)
- ☐ I have included any code used to process the data (or given version details of software used)

Filenames:

- ☐ My filenames are consistent
- ☐ My filenames are unique
- ☐ My filenames are meaningful

README file (project-level):

- ☐ My README is a plain text (.txt) file titled 'README'
- ☐ I have given a clear overview of the research
- ☐ I have given an overview of the types of data included
- ☐ I have listed the file structure and explained the naming conventions for my data
- ☐ I have explained any acronyms or terminology needed to understand the file and folder names
- ☐ I have referenced any third-party data used

README file (folder-level):

- ☐ My README is a plain text (.txt) file with 'README' in the name
- ☐ I have described the methodology for collecting and processing my data
- ☐ I have described how the data is structured within the files (including units)
- ☐ I have listed the file formats used and any specific software needed to open the data

Licence:

- ☐ I have chosen a commonly used licence
- ☐ My licence is as open as possible

Other considerations:

- ☐ I have chosen a FAIR repository
- ☐ I have checked for any additional requirements from my funding
- ☐ I have used widely used terminology (from subject specific standards where possible)

Congratulations, your data is now FAIR! You can share your data and cite it in your publications!

¹Version 2.2: July 2022

²If these apply to you, you are probably already aware of these – see the 'national security concerns' FAQ if unsure

FAQs

What actually is FAIR?

FAIR is a set of principles for data management. The four main principles are that data should be Findable, Accessible, Interoperable and Reusable. These four overarching principles are supported by 15 others, focused on practical implementation. Underpinning FAIR is the idea that data should be easy to find, access and reuse both by humans and automated systems. With this machine interaction in mind, a lot of the focus of FAIR is on the metadata which supports the data, and the structures these both take. The original paper can be found [here](#)³, while a more accessible overview can be found at www.go-fair.org/fair-principles/

How do I choose a FAIR repository?

If in doubt, the University repository ORDA (<https://orda.shef.ac.uk/>), built on Figshare, is FAIR compliant. In general, there are two types of repository that make sense to use:

Discipline specific repositories tend to be smaller and geared towards specific types of data (for example [SimTK](#)⁴ for biomedical computation research). If one of these is relevant to you, you're likely already aware of it; if you aren't already, then consider whether it actually is a good place for others to find your data.

Generic repositories (including ORDA) support many types of data. Sometimes a funder will ask you to use a specific one of these.

In both cases, check whether the repository is FAIR compliant. At a bare minimum, it should provide a unique permanent identifier, support metadata even when the data is not available, and be searchable.

I have chosen a repository that isn't ORDA, what do I need to check?

If a repository has made themselves FAIR compliant, they should make this clear in their documentation. Check that the repository creates a permanent identifier (ideally a DOI), allows for metadata and licences, allows for data storage for at least ten years, and matches your funder requirements. It's also worth checking that any costs are covered by your funders.

What if I'm already using GitHub?

If you've been following best practice (including metadata, READMEs and using accessible file types), then your data is likely most of the way to being FAIR already. However, data needs a unique persistent identifier (i.e. a DOI) to be considered FAIR. A GitHub.com link is not sufficient as, if GitHub shut down or migrated platforms, it would no longer point to your data.

Within ORDA, you can link data from GitHub on the My Data page using the icon below. Alternatively, there is a [GitHub guide for using Zenodo](#)⁵.



Import from GitHub

³Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

⁴<https://simtk.org/>

⁵<https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content>

How do I choose a licence?

Licences should be clearly understandable and allow for automated searches. The license you choose for your data does not have to be the same as any accompanying publications. You should choose your licence to be as open as possible but as restrictive as necessary. For example [creative commons licences](#)⁶ range from CC-BY (most open) to NonCommercial-NoDerivs CC BY-NC-ND (most restrictive), the same principles apply for [software licences](#)⁷.

Do I need to share ALL my data?

No. Not all data has equal value and merits the effort and resources required to preserve. Equally, archiving 'junk' data makes the useful data harder to find. When selecting which data to archive, consider the reasons why it may be needed. For research data, this is often either for reproducibility or because the data set has scientific or historical value, however there may be other reasons. The Digital Curation Centre has a [guide to data selection](#)⁸.

If you are unable to share any of your data, there are still options within ORDA. You can create a metadata only record (without uploading any data), with details of what your data includes and details of who to contact to gain access. There is also the option to add temporary or permanent embargoes to your data and/or metadata, allowing them to become available after a specified time period. More information can be found on the University website at <https://www.sheffield.ac.uk/library/rdm/publish>

My data is saved as a proprietary file type, what do I do?

While it is best practice to have data in an open format, the only requirement from the perspective of FAIR is that it meets community standards. If a file format is not open but is commonly used within your area of research, it can be acceptable. Otherwise, it's often best to convert the data to .txt or similar, if possible.

In terms of best practice, the UK Data Service suggests a [list of recommended file formats](#)⁹ for different types of data.

I processed my data using Matlab (or other proprietary software), do I need to rewrite my code in Python?

No. While open software would be ideal, if your code is in a commonly used language in your field it is acceptable to share as is. With Matlab, the code can still be opened as a .txt file without a licence, allowing the end user to understand it even if they can't run it. However, if your data is saved within a Matlab structure, it might be best to make a copy in an open format.

The data is not mine to share, what should I do?

If you're working with an already available public dataset, then provide links to that. There's no need to reduplicate effort. If you've been provided data by a company, then the handling of this data should be part of your agreement with them ahead of time. Even if sharing all the data is not an option, consider partial datasets or embargos.

What if my data is personal or confidential?

If you're handling personal data, this should be anonymised where possible. You should agree on how you plan to share your data with participants ahead of time and act accordingly.

⁶<https://creativecommons.org/about/cclicenses/>

⁷<https://choosealicense.com/licenses/>

⁸Harvey, R. (2008). "Appraisal and Selection". DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3325. Available online: <https://www.dcc.ac.uk/guidance/briefing-papers/introduction-curation>

⁹<https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/>

What if my data is too big for ORDA?

The library has other options for larger datasets, including group level storage. If these still aren't sufficient, consider sharing partial datasets where necessary. Any questions regarding ORDA can be directed to rdm@sheffield.ac.uk

Will this fulfil my funder requirements?

Hopefully yes, but you should check here: <https://www.sheffield.ac.uk/library/rdm/funders>

What if I don't feel that 'x' belongs in the project/folder level README?

The examples given are meant as guidelines only. If you feel your data is better described by moving things around, then feel free, as long as it's clear for the end user.

My data could have some national security concerns, what should I do?

If you are working in these areas, you're likely already aware of the restrictions on sharing your data (namely export control legislation). Information on export controls can be found at <https://students.sheffield.ac.uk/research-services/export-control>

My data corresponds to a journal publication, do I need to rewrite all my methodology in the README files?

You should aim for your data to be self-sufficient. End users should be able to use your data even in the event that the journal goes out of print. Note that the methodology in the README just needs to be sufficient for people to reuse your data and might be less detailed than you would put in a journal or thesis.

Why should I bother with any of this?

Beyond any funder requirements, making your data FAIR makes it easier for you, your colleagues and other researchers to use your data. Having your data well organised will make your life easier while writing up or publishing and the earlier you consider this, the easier it is. Equally, having self-supporting, easily understood data means that any collaborative work you end up doing with other researchers becomes that much less onerous to set up.

I am still confused

The University's research data management team can be contacted at rdm@sheffield.ac.uk, and are happy to help you figure out what works best for your specific circumstances.

More general information about research data management can be found on the University website at <https://www.sheffield.ac.uk/library/rdm>