**MAKING YOUR DATA FAIR – Checklist (PEP research group, School of Architecture):**

To make your research data and/or software **F**indable, **A**ccessible, **I**nteroperable and **R**eusable, aim to make them ‘as open as possible, as closed as necessary’. *This document provides general guidance for researchers in the PEP research group in Architecture on improving the ‘FAIR-ness’ of data and software.*

**Data Storage and File Structure Management**

*This should be considered at the beginning of the research project*...

|  |  |
| --- | --- |
| Store your digital data securely, preferably in theUniversity’s research data storage or University Google drive. If you collect or create physical data, digitise them if possible. Find more guidance on research storage [here](https://students.sheffield.ac.uk/it-services/research/storage/standard).  |  ⬜ |
| Data Backup: Create and check backups of the data on a regular basis. Use the 3-2-1 principle – always have at least 3 copies, on at least 2 different media, with at least 1 off-site. See above guidance on storage.  |  ⬜ |
| Logical Folder and File Structure:  Create a logical file structure and provide a brief overview of the directory structure and description of the contents of major folders in your readme file (see below). See the [UK Data Service Guidelines](https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/organising/) for advice. |  ⬜ |
| [File Naming Conventions](https://www.sheffield.ac.uk/library/rdm/organising#tab01): Make file names unique, including the most important identifying information of the project. A good file name may include project name, study title, location, researcher initials, and date format, (e.g. YYYYMMDD). It is also important to identify and distinguish versions of research data files consistently. Possible template: ordinal numbers (v1, v2, v3, etc.) for major version changes and decimals for minor changes (v1, v1.1, v2.6). |  ⬜ |
| Research Ethics: If your research involves human participants, make sure that [consent](https://www.sheffield.ac.uk/research-services/ethics-integrity/policy) includes your plans for long-term storage and sharing of data.If you make data available openly or on request, make sure that you are within the terms of participant consent, partner agreements, and third-party data provider agreements. You may be able to share analysed data instead of raw data, but you need to ensure that it complies with stakeholder agreements. |  ⬜ |
| Check if you are permitted to share third-party data, in whole or part, or data derived from them (or the level of granularity at which this is permitted). Don’t assume you can share data just because they are available online. Discuss options for data sharing with external research partners. | ⬜ |

**Sharing Datasets/Files**

Share data that validates your research, especially if it has potential for reuse.

|  |  |  |
| --- | --- | --- |
| Identify an open and trusted [repository](https://www.sheffield.ac.uk/library/rdm/repositories#tab00) for sharing the data. This could be a subject-specific repository (e.g. [UK Data Archive](https://www.data-archive.ac.uk/) (ESRC-funded), [Environmental Data Service](https://nerc.ukri.org/research/sites/environmental-data-service-eds/) (NERC)), one specialising in sensitive data (e.g. [UK Data Archive](https://www.data-archive.ac.uk/)) or a repository for general research data (e.g. the University of Sheffield repository, [ORDA](https://orda.shef.ac.uk/)). Other examples include [Zenodo](https://zenodo.org/) and the [Open Science Framework](https://osf.io/4znzp/wiki/home/). For additional examples of repositories, check the [Registry of Research Data Repositories.](http://www.re3data.org) The University’s website provides additional [guidance](https://www.sheffield.ac.uk/polopoly_fs/1.972501%21/file/Uploading-items-to-ORDA.pdf) and [tips](https://www.sheffield.ac.uk/polopoly_fs/1.972502%21/file/Uploading-items-to-ORDA_Useful-tips.pdf) on how to use ORDA.  |  ⬜ | Findable / Accessible  |
| Assign a globally unique, persistent, resolvable identifier for the dataset (e.g. a DOI). Repositories such as ORDA will assign a DOI upon deposit. |  ⬜ | Findable |
| Datasets/Files Format: Transfer your data to a standard and/or commonly [accessible format](https://howtofair.dk/how-to-fair/file-formats/) if they are in a specialised/proprietary format. (See the [UK Data Service](https://ukdataservice.ac.uk/learning-hub/research-data-management/#format-your-data) website for more information.) Check the University of Cambridge Guidance on best formats for preservation: [www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats](http://www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats)  |  ⬜ | Accessible / Interoperable  |
| Licensing: Documentation should include a description of standard licences applied to the data, and any additional terms of use (for example, [CC0](https://creativecommons.org/share-your-work/public-domain/cc0/) and [CC-BY](https://creativecommons.org/licenses/by/4.0/) for data; MIT and Apache for code). Guides on selecting an appropriate licence: Data: [https://creativecommons.org/choose/#](https://creativecommons.org/choose/) Software and Code: <https://choosealicense.com/licenses/>.  |  ⬜ | Reusable  |
| Secondary datasets: Where possible, if you have used data in your research that is publicly and permanently available (e.g. census data), share a link rather than sharing the actual data.  |  ⬜ | Findable / Accessible |
| Data Archiving and Preservation: Consider what should be kept, what format to keep it in, and where to keep it. Most funders require that data be kept for at least 10 years. Check <https://www.data.cam.ac.uk/funders> for research data policies of major university research funders.  |  ⬜ | Findable / Accessible  |

**README/Metadata**

*If you have data, then you have metadata: these are the who, what, when, where, why, how of your research.* *Rich metadata provides important context for the interpretation of your data and makes it easier for machines to conduct automated analysis.*

|  |  |  |
| --- | --- | --- |
| Follow standard metadata schemes or discipline-specific metadata Schemes. For example, the [Dublin Core Metadata Element Set](https://www.dublincore.org/specifications/dublin-core/dces/) the [Data Curation Centre](https://www.dcc.ac.uk/guidance/standards/metadata) ,the [Metadata Standards Directory](http://rd-alliance.github.io/metadata-directory/) and the [Schema.org](https://schema.org/Dataset).  | ⬜ | Findable / Accessible |
| Make sure the metadata is open access, even if the data has restricted access. | ⬜ | Accessible |
| Does the metadata provide explicit information about all associated data files and their file types, software requirements, and/or conversions? |  ⬜ | Reusable  |
| Is the metadata accompanied by useful terms and notations pertaining to the disciplinary field? (For example, units, common domain identifiers, and spelt-out acronyms) |  ⬜ | Interoperable  |
| In the metadata, are related articles linked and referenced? |  ⬜ | Reusable  |
| If possible, can the metadata be converted into a machine-readable structured text format? (For example, XML or .txt).  | ⬜ | Interoperable  |
| Create clear documentation for your data, ideally in a [README file](https://www.sheffield.ac.uk/library/rdm/organising#tab03). Include your methodology, file structure and any other information that will help you and others to understand and use the data. You can download a README Template here: <https://data.research.cornell.edu/content/readme>.  | ⬜ | Reusable  |

**Preparation of Datasets/Files/Codes for Upload**

|  |  |  |
| --- | --- | --- |
| Group data and files into meaningful datasets. This could be based on an experiment, the processing or analysis method (e.g. simulation models), or a temporal phase.  |  ⬜ | Findable  |
| **For Tabulated Data / Spreadsheets:** |  |
| Do not include special characters and spaces in column headers of tabular data. It is important to include units in the column headers if possible.  |  ⬜ | Interoperable  |
| Do not include charts or images in the spreadsheets. Create separate files for them.  |  ⬜ | Interoperable |
| Save tabs into separate spreadsheets as CSV files. Save all calculations and formulas that need to be preserved as a separate Excel Book (.xlsx), in addition to the original CSV file.  |  ⬜ | Interoperable |
| **For Code:**  |  |
| Make sure your README file includes details of version and dependencies, information about the software environment, and packages required for the code to run. Tools such as **Doxygen** (<https://doxygen.nl/>) can help with documenting your code, including generating UML diagrams. |  ⬜ | Reusable  |
| Include comments in the code in a way that makes it easy for others to understand what is going on.  | ⬜ | Reusable  |
| Include links/paths to other files or directories in the documentation; this includes data files needed for the code to run.  | ⬜ | Reusable |
| Where it is necessary to use a proprietary file format, include details in your readme file of the software and version used. Example: .edt file extension of the ENVI-MET Microclimate Modelling tool. | ⬜ | Interoperable |
| **Sensitive Data and Physical Data** |
| Make access arrangements for physical data if they are important for validation/reproduction of your research and cannot be digitised. Check <https://ukdataservice.ac.uk/learning-hub/research-data-management/store-your-data/security/> on guidance for physical data security.  | ⬜ | Accessible/ Reusable |
| Where not all of your data can be shared, consider sharing some of your data openly through a repository, e.g. fully anonymised, analysed or sample data.  | ⬜ | Accessible / Reusable |

**Remember to…**

|  |  |
| --- | --- |
| Place a [data availability statement](https://www.sheffield.ac.uk/library/rdm/publish) in theses and publications, including a DOI for your data where possible, or contact details for access requests. |  ⬜ |
| Address any confidentiality or privacy issues that may exist.Ensure that you have eliminated or de-identified any data that could be used to identify the subjects of the research. |  ⬜ |
| Make any delay to the release of data as short as possible and within funder requirements  |  ⬜ |
| Check whether individual repositories charge costs for depositing data and whether these will be covered by your funding. |  ⬜ |

For more help and support with research data/software management issues, see the University Library’s guidance [here](https://www.sheffield.ac.uk/library/rdm/publish) or contact rdm@sheffield.ac.uk for support. You can learn more about the FAIR principles <https://www.go-fair.org/fair-principles/>.