

MOORE ABOUT.. README FILES



BETTY AND GORDON MOORE LIBRARY

README FILES

What is a README file?



With README files, the clue is in the name! It's a plain text file included with a dataset which potential users are encouraged to read. It helps to explain the contents, format and context of any data so that others can make the best use of them possible. You will have spent many hours with your data and have a complete understanding of what things mean but new users won't have this privilege. Think of a README file as a map to guide these potential users through your data. They may recognise individual elements but need to understand how things work together to fully interpret the data.

Why are they important?

As the world increasingly embraces open research practices, individual researchers are encouraged to share their data with others. Not only does this help to reproduce and verify results but it also means that other researchers can build upon existing work and not spend time replicating work that has already been done. In order to make the best use of these data it's important that researchers understand them and this is where a README file is crucial. The [FAIR Data Principles](#) advocate that not only should data be discoverable but also reusable and this is a key factor in the creation of a README file.



Putting together a file also forces the creator of the data to think carefully about how they present these data – are columns labelled appropriately? Are known standards being used and if so which ones? Are appropriate licences attached to the data? Getting into good data management and organisation habits early in the research process will make the creation of shareable, discoverable, accessible data easier for all.

AREAS TO COVER

Below you will find some of the most common areas covered by README files. As there is no standard format you should always check with your funder/publisher to ensure that you have included all of the elements they require. If you have freedom to design your own file then think about which areas you would find most helpful as a potential user of the data.



PROJECT DETAILS

Record the name of the project, the year(s) it took place and name/contact of the lead author(s). You should also include an acknowledgement of any funding which has supported the project. If you have a DOI for the data or any related publications you can also include these.



PROJECT SUMMARY

Write a short summary of your project including brief aims and outcomes. Include subject specific keywords that describe your project.



CONTACT INFORMATION

Include the name of the Principal Investigator or the person with ultimate responsibility for data collection together with their affiliation and contact details. Remember that you may want to include alternative contact details in addition to institutional emails in case of movement.



DATA SUMMARY

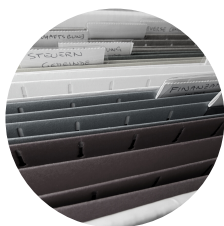
Write a brief summary of your data such as how and when it was collected and how it was processed, including if you used any specialist software.



FILE NAMING CONVENTION

Decide on and detail the file naming convention you have used for your data. Remember that it is important for any scheme used to be consistently applied, meaningful and easy to understand so avoid over-long file names.

AREAS TO COVER



FILE FORMATS

Detail which format(s) have been used to share data, including whether these are proprietary or not. Try and include accessible formats such as PDFs and open source text documents if possible.



METADATA STANDARDS

Include information about the metadata standards you have used and how these were applied.



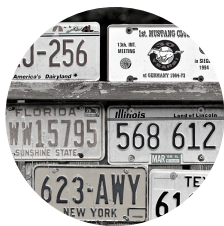
CODES

If you have used any codes or abbreviations, you should clearly outline what these are and how they work. Remember that although they may make sense to you, other researchers will lack the same context.



FILE LIST

Give clear information on each file in your dataset including its name (using any file naming convention), a brief description of its contents, any specific element of the research it refers to (e.g. survey question number), the date it was created and the file format (if not included elsewhere).



LICENCE INFORMATION

Think about any licences that you want to attach to individual files or the dataset as a whole. You may be required to use a certain licence or you may have the freedom to select one. Consider an open licence such as [Creative Commons](#) if appropriate and remember that you can only attach licences to materials you own the copyright to.

TRY IT OUT

Want to put together a README file for your project? Use our [interactive online tool](#) to try it out for yourself. And don't forget to take a look at our [Data Management guide](#) for more information on how to organise your data throughout your project.



This work is licensed under a Creative Commons CC-BY 4.0 license by Cambridge University Libraries.