

Messy data can limit your research analysis and opportunities for further use. By adopting a logical, consistent approach to how you organise and label research data, you can gain a better understanding of the data that you hold and that which are missing or required; avoid confusion when working with multiple versions of the same work; save time when locating resources; and avoid the storage of duplicate copies of the same data

Label files and variables consistently

- **Adopt a naming convention:** Use standards where they exist. E.g. CDISC for clinical research.
- **Be descriptive:** Filenames and variable labels should be sufficiently descriptive to understand their purpose/content. E.g. StudyProtocol.docx, Codebook.docx.
- **Avoid spaces:** Many software applications are confused by spaces in filenames & variables. Replace these spaces with hyphens (-) or underscores (_). E.g. 'wales survey results.csv' could be renamed 'wales-survey-results.csv'
- **Use upper/lower case consistently:** Linux & Mac OS allow you to label two files with the same name, where the only difference is capitalisation (e.g. FILENAME.PDF and filename.pdf). Microsoft Windows treats both files as the same, however, and will overwrite one of them.
- **Don't use reserved characters:** Some characters perform specific functions in operating systems and software (e.g. forward/back slash, colon, quotes) and can only be used for other purposes with care. See http://en.wikipedia.org/wiki/Filename#Reserved_characters_and_words.
- **Label filenames to make them easy to sort:**
 - **To sort files chronologically:** Place the date (YYYY-MM-DD) first, e.g. 20130912_Agenda.doc.
 - **To sort by participant:** place the participant ID first, e.g. 021Interview.doc
 - **To sort sequentially:** Use a number sequence, e.g. WorkPackage01, WorkPackage10.
- **Label each version consistently:** For example d1, d2, v1, v2. This is particularly important when storing the same content in multiple formats, e.g. '1234-transcript-v1.1.docx' and '1234-transcript-v1.1.odt'.

Book an appointment

To discuss your data management approach, contact the LSHTM Research Data Management Service at researchdatamanagement@lshtm.ac.uk.

Further information:

- Centre for Environment Archive: Filenames for science data
<http://web.archive.org/web/20151217045723/http://www.ceda.ac.uk/help/users-guide/filenames/>
- MANTRA Online Learning Module: Organising Data <http://datalib.edina.ac.uk/mantra/>
- UK Data Service: Organising data <http://ukdataservice.ac.uk/manage-data/format/organising.aspx>
- JISC Digital Media: Choosing a file name
<http://web.archive.org/web/20160414224118/http://www.jiscdigitalmedia.ac.uk/guide/choosing-a-file-name>