How to prepare your research data for deposit at LSE

To deposit your data files in the LSE data repository, you will need to prepare it for ingestion and ensure its suitable for archiving and sharing. A deposit can be delayed or rejected if it's not prepared properly, for example, you have not cleared intellectual rights in a dataset or anonymised a dataset where this is required. For further help, contact us at datalibrary@lse.ac.uk.

1 – Check your data

- Check your dataset falls under our collection policy and that you are happy with the terms of the LSE repository deposit licence.
- If your dataset includes participant data, check that you have documented consent to sharing data. Even where participant data has been collected fully anonymised, it is good practice to inform participants of your intention to share data. Consent should be specific - general statements such as 'data may be shared with other researchers' is not sufficient. You should not disclose data if your consent procedure stated that data would not be disclosed or destroyed. See LSE informed consent guidance for sample consent forms including data sharing.
- If your data includes any third-party copyrighted or restricted materials, you must ensure you have either the rights to share the data, seek permission to share, or remove the material from the deposit. Where datasets include secondary data, the original data owners may be able to say how and on what terms their data are distributed by you. You may need to check any applicable research contracts to establish what parties hold rights in a dataset. For further information on copyright please see LSE copyright hub.
- Check the size of your dataset and ensure it does not exceed the repository size limit. This allows up to 20 GB freely, with additional storage being available at a cost.

2 – Organise your data

- Please ensure all the data are contained in a zipped folder - this enables files transferred into the repository system to retain their underlying folder structure.
- Within your top-level folder, you must also contain a file entitled 'readme.txt'. The contents of this file are described in the next section.
- Make sure your data folders are organised in a logical fashion. Try to think about how a user of the data would need the data to be structured for use.

- Wherever possible, please ensure the data are stored in suitable formats for preservation, for example by using open rather than proprietary file formats, eg, csv instead of Excel.
- Give files meaningful and consistent names and ensure relevant information is clearly presented in files, eg, variable names, units of measurement, missing value codes etc.
- Redact data as necessary for anonymisation or commercial purposes. There is guidance on anonymisation provided by the UK Data Service.

3 – Document your data

- Your deposit must include a readme file, which should, as a minimum, include:
  - A file inventory of the dataset
  - References to any secondary data sources used
  - Details of any particular operating system and/or software required, or any other dependencies to make use of the data
  - key interpretative information, eg, descriptions of column headings and row labels, full definition of variables and units used etc.
- Your dataset should also consist of documentation and metadata. This may include a user guide, codebook/ data dictionary, copies of interview schedules or questionnaire templates, experimental protocols etc.
- Where relevant, a blank consent form and a copy of the patient/participant information sheet must also be included. These will be stored for administrative purposes only. Access to them will be restricted, meaning they will not form part of the dataset available for users download.