Protocol on the protection, anonymisation and sharing of research data

This protocol is designed to give researchers a framework for handling personal and confidential data and sharing research data within the context of the Data Protection Act 2018 (DPA) and the ESRC Framework for Research Ethics. It has been developed from guidance issued by the UK Data Archive (UKDA) and the ESRC.

For further information on how to share personal and confidential data securely please also see the University's Information Protection Policy.

Handling personal data

1. It is important that those undertaking research are aware that most of the Data Protection Principles embodied in the DPA apply to their work. This means that 'personal data', defined as data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller:

   - must be obtained for a specified and lawful purpose;
   - shall not be processed in any manner incompatible with that purpose;
   - shall be adequate, relevant and not excessive for those purposes
   - shall be kept up to date;
   - shall be kept for no longer than is necessary for that purpose;
   - must be processed in accordance with the data subject's rights
   - must be kept safe from unauthorised access, accidental loss or destruction and;
   - shall not be transferred to a country outside the European Economic Area unless that country has equivalent levels of protection for personal data.

2. Personal data, such as informants' names, addresses, etc, should only be collected if necessary for research purposes or follow-on research. Often such data are collected for administrative purposes only and have no research value. Not collecting personal data in the first place may make it easier to manage data. If data does need to be collected, for example, for follow-up interviews, they should be stored separately from research data.

3. Researchers also have a duty to collect and use special category or confidential information appropriately. Special category personal data are defined in the DPA as data on a person's race, ethnic origin, political opinion, religious or similar beliefs, trade union membership, physical or mental health or condition, or sexual life. Definitions of confidential and highly confidential data can be found in the University's Information Protection Policy.

4. Note here that the DPA applies only to personal data. This means that once data have been anonymised, the DPA no longer applies. Please be aware of the distinction between genuinely anonymised data, (it cannot be traced back to an individual) and pseudonymised data (defined in GDPR as "…the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.")

Gaining consent for the use of personal data

5. Whilst researchers have an ethical responsibility to secure the consent of participants before their personal data is collected and used, under DPA **public task** is the lawful basis for data processing for research purposes. This means that:

   - The data can be retained indefinitely (if needed).
   - The data can be used for future research or deposited in an archive. (In most circumstances this would be anonymised data anyway.)

6. This means that, whilst it might not be possible to provide detail on how their data *might* be used, participants should be asked to provide consent for the *possible* uses of their personal data at the outset of the project.

   - The Research and Innovation Service website has guidance on obtaining participants' informed consent, including template information sheets and consent forms.
   - Please direct participants to the University's Research Privacy Notice.

Online survey tools

7. It is good practice to use Online Surveys for all online surveys because most similar tools store data outside of the European Economic Area and so may not comply with the DPA; it is free for University of Leeds researchers to use, more robust and has more features than many of the online tools.

Cloud computing services for storing and accessing data

8. Many cloud computing services such as Apple iCloud, Dropbox, Amazon S3, Google Apps etc. store data outside of the European Economic Area and thus may not comply with the DPA. The terms of use normally state that the provider accepts no liability for temporary or permanent loss of data. For these reasons, the University advises that the only acceptable cloud computing services used to process data are those provided by the institution such as Office 365. If you have any concerns about this please contact your faculty IT Manager for further assistance. Please also see the Information Protection Policy.

Anonymising research data

9. Before research containing data about people can be disseminated, made public, archived or shared with other researchers consideration should be given as to whether it can be anonymised.

10. Anonymisation is needed for ethical reasons to protect people's identities in research, for legal reasons to not disclose personal data, or for commercial reasons. Anonymisation should cover the removal of 'direct' and 'indirect' identifiers:

    ○ **Direct identifiers**
    These include names, addresses, postcode information, telephone number etc. and are often collected as part of the research administration process but are usually not essential research information.

    ○ **Indirect identifiers**
    These include information that, when linked with other publicly available sources, could result in a breach of confidentiality. This could include geographical information, information on workplace, organisation, education institution or occupation.

- **Anonymising quantitative data**
  Where relations between variables in related datasets can disclose identities, and where identifying geo-references also have a geographical importance.

- **Anonymising qualitative data**
  Where data are destined to be archived and made available for sharing, it is important to reach an appropriate level of anonymity, whilst aiming to maintain maximum meaningful information in the research data. In a textual version of data, information should not be removed or blanked-out, but rather pseudonyms, replacement terms or vaguer descriptors should be used.

11. Similar concerns apply to non-textual data such as digital images and audio or video recordings. Again, editing to remove identifying detail should be done sensitively. A word here or there in an audio recording may be bleeped out (for example to remove people's names). Whilst it is technically possible to disguise voices by raising the pitch in a recording, or to obscure faces by pixellating sections of a video image, such approaches significantly reduce the usefulness of such data. Pre-planning and agreeing with participants during the consent process on what may and may not be recorded, will be much more effective in creating data that accurately represent the research process and the contribution of participants.

12. Also, given the ability to match data from other sources, it is important to consider not just the data you hold but the data environment that it will interact with in order to determine whether it has been genuinely anonymised. The anonymisation network has established a 10 point anonymisation framework that can be used to assess this.

13. The Research and Innovations Service website has further guidance on confidentiality and annonymisation.

Sharing data

14. Archiving and secondary use of research data is becoming more common and is considered good practice by many funders and professional societies. For example the ESRC Framework for Research Ethics states *"Researchers who collect the data initially should be aware that ESRC expects that others will also use it, so consent should be obtained on this basis and the original researcher must take into account the long-term use and preservation of data."*

15. It is therefore important that researchers address data sharing early in research planning and as part of the consent process, so that measures can be put in place to safeguard participants and the information they provide and to obtain appropriate consent for a variety of data uses. You might need to have a Data Processing or a Data Sharing Agreement in place to protect the personal data that you are sharing; templates are available on the University's Data Protection website. If you are sharing the data with an organisation outside the EEA you can use still the templates if the receiving party agrees. If this is not appropriate then you will need to reach a common, documented understanding on how personal data will be securely shared, stored and deleted; or else only share anonymised data.

Explaining data sharing to participants

16. Researchers should inform all participants about how their data may be processed and stored in an archive, and give them the opportunity to consent or dissent.

17. Care should be taken to explain to participants the value of archiving research data, and of making it available for future use. Detail should also be provided on what kind of archive their data might be put into, and who might be able to access it. Typically this would mean that the data would not be in the open public domain but available for research and education purposes with the undertaking that other users would:

- Not disseminate any identifying or confidential information on individuals, households or organisations.
- Not use the data to attempt to obtain information relating specifically to an identifiable individual.

18. Arrangements for archiving special category and confidential data arrangements are more complex. This data can be shared ethically if researchers apply one or more of the following strategies:

    - Obtain informed consent for data sharing, as well as obtaining consent for participation and for other uses, such as publication.
    - Protect people's identities when necessary, by anonymising research data
    - Decide if access restrictions to all or part of the data may be required.
    - For especially sensitive data additional restrictions could also be applied such as:
        - ➢ specific data access authorisation from the researcher prior to release of the data
        - ➢ an embargo on further use of the data for a given period of time until confidentiality is no longer pertinent.

    These procedures should always be considered jointly, not in isolation, and researchers should discuss them openly with participants.

Assuring ethical re-use of data

19. Putting data into an archive is not the same as making them available on the web. Archivists value the materials deposited with them and take their duty very seriously to make sure the materials are used only in appropriate ways. Their primary concern is to protect research participants.

20. Archives, such as Research Data Leeds, or the UKDA, use licences to control access to the data. Some data are available to the public, some are covered by a standard licence, some need special permission, and some data are made unavailable for a lengthy period.

Further guidance

Information Protection Policy

The University's data protection website

Research data management explained