Research Ethics Committee DMP Guidance

Introduction

This guidance accompanies the REC DMP Template. The template should be used when making an application for ethical approval to the University Research Ethics Committee (UREC) or a School Research Ethics Committee (SREC) that requires submission of a DMP as part of an application for ethical approval.

The REC DMP Template is for use with REC submissions only. It is not a general research project data management planning tool. The RDM website provides guidance on general-purpose DMP tools.

You should complete the DMP with reference to the guidance below and the Data Protection for Researchers guide provided by the Information Management and Policy Services (IMPS) office. You can also refer to the REC DMP Examples document, which provides examples of suitable responses for inclusion in the DMP.

All sections of the DMP must be completed; if a section is not applicable, enter N/A. For submissions to UREC, the DMP will be reviewed by the IMPS office and the Research Data Manager, and feedback will be included with the UREC opinion. Any SREC that requires a DMP will review it as part of the ethical review and provide feedback with its opinion.

If you have questions about completing the DMP, please contact Robert Darby, Research Data Manager, r.m.darby@reading.ac.uk / 0118 378 6161.

Definitions

The DMP is concerned with how you will manage both personal data and confidential information in relation to research data.

Personal data is information as defined in data protection laws relating to natural persons who can be identified or are identifiable, either directly from the information, or indirectly from the information in combination with other information. Personal data must be processed in accordance with data protection laws. Personal data that contain sensitive information, such as details of a person's medical history, constitute a special category under data protection laws, and are subject to stricter processing requirements.

Confidential information is non-public information relating to an identifiable living individual or legal entity. Those with whom this information is shared have an ethical obligation to maintain the confidentiality of the information. In research, confidential individual information and personal data are often the same. Information may be confidential for other reasons, for example, if it is non-public information relating to a

business or other organisation. Examples include commercially confidential information, and sensitive location details (e.g. relating to endangered species, military sites, etc.).

Research data is information collected to answer a research question. Where the data have been collected from research participants, they are usually processed at some stage in the research to remove direct identifiers such as names and contact details, although they may still include indirect identifiers (such as key codes, age, job title, etc.). De-identified research data is generally suitable for public sharing, although care may need to be taken to unlink pseudonymous key codes and mask or remove indirect identifiers before making research data widely available.

Anonymous data is data that no-one can relate to an identifiable individual. For example, if a survey did not record any information relating to the identity of an individual, including data such as an email address or telephone number, then the data collected could be described as anonymous, providing that the information itself was not sufficiently detailed enough to *indirectly* identify them.

Pseudonymous/de-identified/link-coded data is data that relates to an individual but is presented in such a way that it does not directly identify them. If the researcher still knows who the data relates to, it is still personal data in data protection law. An example is where a unique participant ID is applied to a dataset, and direct identifiers (such as name and contact details) are removed. These identifiers may be held in a separate link table that lists identifiable participants and the linked IDs by which their data records are coded in the dataset. Separating identifying information from research data during research is a recommended data minimisation practice because it reduces the risk that unauthorised persons could identify data subjects if the dataset is compromised. But under data protection law pseudonymous or link-coded data remain personal data and cannot be described as anonymous.

Anonymised data is data from which identifiers (both direct and indirect, and including linked ID codes) have been removed, such that the data cannot be related to an identifiable individual. There is a body of opinion that it is increasingly hard in the digital world to say with absolute confidence that data could not be reverse-engineered or re-identified by linkage with other data. But researchers can still give assurances to participants that data will be anonymised. Attempting to re-identify anonymised data is also an offence in data protection law, which thereby offers some safeguard for published anonymised data. Researchers should nevertheless carefully scrutinise data prepared for disclosure, and be alive to the possibility that some data may have the potential to be indirect identifiers when combined with other sources of data that may be available to some people. Guidance on anonymisation techniques for both quantitative and qualitative data can be found in Data Protection for Researchers.

Guidance on completing the DMP

1. What research data will be collected?

Describe:

- the types of data to be collected, e.g. interviews with farm workers about pesticide use; analyses of blood samples from trial patients; structural and functional MRI scans, with processed images, statistical data and analysis code;
- all media/formats data will be collected in, e.g. paper questionnaires; audio recordings; text transcripts; online tools with export to standard tabular formats; proprietary instrument formats;
- the anticipated quantity or scale of each type of data, e.g. 20 one-hour interviews; analyses of blood samples taken from 30 patients weekly over 8 weeks; MRI scan sessions of one hour's duration for 50 participants, generating in total ~ 1 TB raw and processed data.

2. What personal data and confidential information will be processed?

2.1 Personal data

Specify the personal data (as defined in data protection laws) that will be collected, e.g. name, address, email address, photographs in which individuals are identifiable, geolocation data, IP addresses.

A checklist of personal data is provided in the DMP.

2.2 Special category data

Specify any special category (sensitive) data you will collect, as defined in data protection laws, e.g. information about an individual's health, political opinions, etc.

A checklist of special category data is provided in the DMP.

2.3 Confidential information

Specify any confidential information not specified above that will be collected, e.g. non-public information relating to a business or other organisation.

Guidance

See the entries for Personal Data and Sensitive Personal Data in Data Protection For Researchers. Personal data may be contained in research data and administrative information relating to the research, such as signed consent forms and correspondence.

3. How will data be stored and transferred during the project?

3.1 Locations

Identify all locations where data and supporting materials will be stored, including:

- the primary storage location for project data and documents, e.g. a location on the University network or a University OneDrive account;
- instruments for data collection and analysis, such as online survey tools, analytic instruments, audio and video recording devices;
- working locations for data processing and analysis, such as project members' laptops and other devices;

- locations for storage of non-digital data, e.g. signed consent forms, paper questionnaires, such as a locked cabinet in a locked office on University premises. Consider storage in the field/in transit as well as on University premises.

For each location, indicate whether it will be used to store/process identifying information or de-identified research data, and provide details of access controls that will be applied, such as password protection, or encryption of files or devices.

3.2 Risk management

Describe any administrative measures that you will take to control the risks of inappropriate disclosure of personal data/confidential information. These might include:

- encrypting any hardware that will be used to store such data (such as audio-recording devices and laptops) and deleting data from these devices as soon as they have been transferred into the primary storage location. Storage on external devices should be avoided as much as possible and should always be temporary;
- digitising hard copy data, including consent forms, for secure digital storage, and destroying paper originals. Consider using digital consent methods where possible: for example, the University's REDCap platform provides an e-consent function;
- storing hard copy data, including consent forms, in a locked cabinet in an office on University premises that is locked when not in use;
- storing participant records separately from research data in folders/areas accessible only to authorised users;
- de-identifying/pseudonymising/link-coding research data by removing direct identifiers and using a unique code to designate each participant that is linked to participants' details in a separate, secure link table;
- establishing protocols for sharing of data within the project team, covering secure transfer between locations as necessary. Use secure end-to-end encrypted methods of transfer wherever possible, e.g. OneDrive or Teams, or VPN connection to the University network; avoid less secure methods, e.g. email, and encrypt files if necessary.

3.3 Authorised persons

Specify who will be able to access the identifying information and how you will ensure they process the information securely, e.g. through training, supervision where appropriate, and adherence to agreed protocols for accessing confidential information in secure environments. If you will be using a service supplier acting as a data processor, such as a professional transcription service, have the terms of service been discussed with and agreed by Procurement or your IT Business Partner?

Guidance

Storage of hard copy personal data in insecure areas such as shared offices that are not locked when unattended or open to the public or visitors must be avoided. Where information contains sensitive data, a principle of two barriers should be applied, such as

using of a locked cabinet within an office that is locked when unattended. Unnecessary duplication of personal data should be avoided. Where possible, signed consent forms should be held in secure electronic storage: paper forms can be scanned and destroyed if there is no good cause for retaining the paper originals.

See the RDM web pages for guidance on data storage and information security, and online survey tools (bottom of page).

Guidance on pseudonymisation can be found in Data Protection for Researchers.

Guidance on data protection requirements for suppliers is provided by Procurement.

4. How will research data be preserved and shared on completion of the project?

Note: This section does not need to be completed by or on behalf of undergraduate or taught postgraduate students, unless there is an expectation that the results of the project will be published. If this section does not apply to the project for which ethical approval is sought, simply enter N/A.

4.1 Research data to be preserved and shared

Identify the research data that will be preserved and shared at the end of the project by deposit in a public data repository. If some or all research data will not be shared, explain why this is the case, for example because they cannot be easily anonymised (as may be the case with video data), or because they are confidential for commercial or other reasons.

4.2 Preparation of data for sharing

Describe the measures that will be taken to ensure data are suitable for sharing, e.g. informing participants during recruitment that data will be shared; securing consent for data sharing; anonymising data prior to deposit/sharing; or depositing confidential data in a data repository under a controlled access policy.

4.3 Data repository

Identify the data repository or repositories that will be used to preserve and share data. If no data repositories will be used, explain what other solutions will be used to ensure research data are preserved and can be accessed publicly or on authorised request after the project.

Guidance

The University's Research Data Management Policy requires researchers to preserve primary data collected in support of published findings, and to make them accessible to others by deposit in a suitable public data repository, unless there is a valid legal, ethical or commercial reason for withholding access to them. The RDM web pages provide guidance on choosing a data repository.

Most research data collected from research participants can be safely and ethically shared once they have been anonymised (i.e. identifying information has been removed), and it is often possible to redact information that is confidential for other reasons. It is not acceptable simply to state that research data cannot be shared for confidential reasons. If you do not intend to share data you must explain why they are not suitable for sharing.

Be aware that in order for publicly-disclosed data to be fully anonymised, any means of linking them to participant records stored internally should be destroyed. If you have used pseudonymous key codes in your dataset which are linked to internal participant records, the key codes in the public dataset should be replaced by random identifiers or the link table destroyed. Guidance on anonymisation can be found in Data Protection for Researchers.

Where data cannot be rendered safe for sharing as open data (for example, where identifying information is intrinsic to the data and cannot be removed, or where data can be anonymised, but may present a higher risk of reidentification through linkage to other information), it may still be possible to share them with authorised users on a restricted basis. Some data repositories provide controlled access procedures for managing safe access to confidential research data. For example:

- Anonymised data can be deposited in the UK Data Service ReShare repository as safeguarded data. These data would only be available in confidence to registered researchers, under the terms of an end-user licence.
- The University's Research Data Archive can offer a restricted dataset option. This is suitable for higher-risk anonymised data or identifiable data. A restricted dataset is held securely by the University and made accessible only if authorised by a Data Access Committee including the PI or nominated data steward and subject to a data access agreement between the University and a recipient organisation (which must be a research-performing organisation with which the researcher requesting access to the data is affiliated).

Further information about restricted archiving services can be found on the Where to archive data web page.

Consent is not required to share anonymised data, although as a matter of good practice research participants should always be informed of plans to make any data collected from them available to others.

You should in any case take care that your consent procedures do not preclude sharing of research data, either as open data or on a restricted basis if necessary. Do not set a time limit on the retention of the data collected from participants, or state that all data will be destroyed at the end of the project, or undertake that data will not be shared outside of the project. Such undertakings are not required by data protection law or research ethics policy, and they will prevent you from making your research data accessible to others, even if they have been anonymised. If you are planning to share data as open data, do not say they will made available to certain groups of users only, e.g. researchers. Open data by their nature can be used by anyone.

A sample consent form, with consent formulae for sharing of anonymised open data, and for restricted sharing of data under safeguards, can be found on the IMPS website. The UK Data Service provides guidance on seeking consent for data sharing, and includes a model consent form.

5. How will retention and disposal of personal data and confidential information after project completion be managed?

5.1 Retention period

State how long you plan to retain personal data/confidential information after the end of the project.

5.2 Responsible person(s)

Specify under whose authority this information will be maintained and disposed of after the project.

Guidance

Personal data should be retained as long as necessary for the specified purpose. It is acceptable under data protection law to retain personal data, used for research purposes, for long periods, subject to periodic review, if they are held for archiving purposes in the public interest, for scientific or historical research purposes, or for statistical purposes. For example, if follow-up studies are contemplated, then continued retention is justified, providing this purpose has been notified to the individual. Care should be taken to avoid commitments to the Research Ethics Committee or participants to destroy personal data by a given time, e.g. 3 years after the completion of the project. It is better to indicate that data will not be held longer than necessary for the specified purpose(s), and schedule regular reviews of personal data holdings to determine whether they need to be retained or can be safely destroyed.

Each research project is unique, and researchers may need to apply a case-by-case assessment to determine how long records and data should be kept. However, it is recommended that consent forms, as a minimum, are retained for the period of time that the research study data is held in identifiable (including coded) form + 5 years. Consent forms can be retained for longer periods if the nature of the research deems it necessary. Exceptions to the minimum retention period will apply where longer retention periods are required in accordance with any contractual, legal, or statutory obligations or directions that are specific to your study, or any terms and conditions imposed by external research sponsors/funders/third parties, for example the Medical Research Council (see guidance on retention periods in the MRC Good Research Practice guide, p. 10).

If personal data/confidential information will be retained in the long term after the completion of the research, planning should take into consideration where and under whose authority they will be held, and what provision is made for transfer of ownership should the original owner(s) leave. It is always advisable for personal or confidential information to be owned by/accessible by more than one person, to avoid the risk of data becoming orphaned when a sole owner leaves the University. For example, a research

group might maintain a personal data asset register, listing personal data held, owners of the data, storage locations, the retention schedule, and the date of next review. The primary owner might be the PI of the original project, with the schedule administrator having access for administrative purposes, and ownership defaulting to the Head of School in the absence or on the departure of the primary owner. This register could be updated on an annual basis and ad hoc when a review date is reached, or when any listed data owner leaves the University.

Guidance on the retention of personal data can be found in Data Protection for Researchers.