MANAGEMENT AND PROTECTION OF DATA COLLECTED FOR RESEARCH PURPOSES

Contents

Introduction

This guidance has been produced to supplement the University's Policy on the Management of Data Supporting Research Outputs[1] and the University's Data Protection and Research guidance and is intended to assist staff and students whose research involves human participants, or personal[2] or special category[3] (previously known as sensitive) data[4] as defined in the General Data Protection Regulation (GDPR).

For information about all aspects of research data management and planning please also see the University's research data website.

For information about the GDPR and its implications for research please see the University's guidance on data management and research.

Informed Consent

Before such research starts, the participants should normally be fully informed about how their data will be managed by the researcher. It should be clear, for example:

- what specific data is being collected (e.g. list personal and special category)
- how it will be gathered/ transferred/ transcribed
- how it will be de-identified (if applicable)
- who will have access to it
- where it will be stored (and for how long)
- what potential use may be made of the data (e.g. sharing with others, publication, use in future research)

Researchers must avoid making promises that may be difficult to keep, e.g. that data will only be seen by the PI, or that all data will be destroyed at the end of the project. It is likely that research data will be seen by research teams and technical/ IT support, so it would be wise not to restrict who may see the data too much in the participant information and consent documents/ scripts unless there are strong reasons for doing so. Equally, the research data should ideally be preserved as long as possible for academic use. There is a minimum storage period of three years after publication according to University policy. Please see 'Retention of data' for further information.

Please also see CUREC's guidance on informed consent, including recommended templates.

---

[1] University of Oxford policy on the management of data supporting research outputs

[2] **Personal data**: data that relate to a living individual who can be identified (a) from those data, or (b) from those data and other information that is in the possession of, or is likely to come into the possession of, the data controller (e.g. through the use of a code devised by, or accessible to, the researchers). Examples include, but are not limited to, name, email address, audio/ video recordings, identification number, IP address, location data, genetic data and biometric data.

[3] **Special category/ sensitive data**: data relating to race, ethnic origin, sexual orientation, political opinions, religious beliefs, physical/ mental health, trade union membership, genetics, sexual life, biometrics (where used for ID purposes), or criminal activities. Special conditions apply to the processing of this type of information.

[4] Note that personal data that has been pseudonymised – e.g. key-coded – can still fall within the scope of the GDPR depending on how difficult it is to attribute the pseudonym to a particular individual.

Safe data gathering and storage

While gathering data in the field, mobile devices containing University data must be protected by whole disc encryption. OneDrive for Business, provided as part of the University's Nexus365 offering, has been approved by the University's Information Security team for the storage of research data. Personal accounts with third party cloud storage providers such as OneDrive, Google Drive, or Dropbox are not generally an appropriate place to store research data (especially sensitive ["special category"] data) unless all data is encrypted, and then only for short-term storage/ transfer of data. Please contact your local departmental IT support or Research Data Oxford (via their email address)for advice on appropriate platforms for storage and sharing of large data sets when OneDrive for Business is not suitable, or in the case of collaboration with third parties outside the University.

Personal and special category (formerly known as sensitive) data must be transferred and then stored as safely and securely as possible, e.g. using encrypted laptops, encrypted USB sticks (only for short-term storage and/ or transfer of data), encrypted files in departmental or centrally-provided storage systems (e.g. SharePoint or Nexus365 OneDrive) or locked filing cabinets.

Researchers must consider the security of the re-transmission of all data if shared with the participant for the purpose of checking the accuracy of a recorded statement. Again, the Research Data Oxford team (via their email address) team can advise on this.

Plans for your research should include a framework that indicates how this will be achieved during and after the research project. Please see Research Data Oxford's guidance on data management for further information.

Anonymisation and identifiers

In general, data should be managed and used in such a way as to protect the confidentiality of the research participants. This is of particular importance if the data involve personal interviews or results from standardised cognitive tests, where the participant would not want results disclosed to others. Anonymisation (or de-identification) is one option (the other is restricting access to data) and needs to be considered in relation to the demands of the project and the expectations of participants. Some base level anonymisation is advised in the handling of data files as well.

For example, it is good practice in general to use a code number to label all paperwork, physical media (e.g. audio recordings, CDs) and computerised records (even discussion via email), with a key giving identities stored separately. The benefits and drawbacks of anonymisation regarding the security and quality of your data need to be considered.

Levels of anonymisation

The level of anonymisation that is necessary will depend on the research. The important point is that if data are not fully de-identified, participants should be aware of this and give their consent. When planning anonymisation measures, consider in what ways identification of participants may occur (i.e. whether a combination of details could allow an individual to be recognised, or if the sample size of participants is relatively small) and what potential consequences this identification would have. A combination of details could allow an individual to be recognised. The aim is to maximise data security and ensure data privacy while minimising the risk of information loss or a data breach. When research participants have been told upfront that the data will be archived for scientific reuse, data need to be anonymised to a level that ensures that re-users of data cannot identify individual participants. Audio-visual datasets cannot be easily anonymised and so can be archived only if explicit consent was given for this.

In most cases the participants' exact names and addresses should be destroyed after the original research has been completed. However, it may be necessary to retain contact information for longitudinal studies; again, the key point is to ensure that the participant has given explicit consent for this.

Where possible, other direct identifiers such as postcodes, telephone numbers, and exact birth dates should be removed from the data after the original research has been completed. Preserving them is justified only when direct identifiers are essential for the analysis of the data, and the participants have given specific consent to the arrangement beforehand.

Access to data

It is not a requirement that access to the research data be limited to the research team only. In fact, promises of restricting data access too much can make it difficult for Oxford to fulfil its responsibilities as data controller. Again, the key point is that if it is planned to make the data more widely available, then the participant must be told of this at the outset.

Data archiving

Given that many research funders are encouraging data archiving, it is a good idea to consider this at an early stage. The UK Data Archive notes: "Consent forms should not preclude data sharing. Promises to destroy the data or that the data will only be seen or accessed by the research team must therefore be avoided. Terms such as 'fully anonymous' or 'strictly confidential' are to be avoided, as they are often impossible to define. Better is to indicate how data will be anonymised or de-identified (e.g. by removing all personal information that could directly identify an individual) and that whilst data will be made available to other researchers, confidentiality will be protected."

For more information, see the UK Data Service's guidance on consent for data sharing.

Data sharing

Research data may be used or disseminated for research purposes only, and the participant's consent to data sharing should be sought. Handing over data to third parties or talking about individual participants to outsiders in a way that would affect the evaluation, treatment, status or behaviour of the participant is unethical. If researchers plan to share personal data or are using a third party (including services such as a transcription service) to collect or process personal data on their behalf (a data processor), they need to seek advice from Research Services to enter into an agreement with that third party to ensure the information is processed in accordance with the University's legal obligations. It may also be appropriate to contact the University's Information Security or Information Compliance teams to check they are satisfied with the provider's information handling practices. Researchers also need to agree with the other party what happens when they no longer need to share the data.[5]

OneDrive for Business, provided as part of the University's Nexus365 offering, has been approved by the University's Information Security team for the sharing of research data. Please contact your local departmental IT support or Research Data Oxford (via their email address) for advice on alternative platforms for storage and sharing of large data sets when OneDrive for Business is not suitable, or in the case of collaboration with third parties outside the University.

Indirect identifiers/ background information

[5] See the ICO's guidance on storage limitation under the GDPR (January 2020)

The following are examples of background variables or indirect identifiers: gender, age, education, occupation, economic activity, socio-economic status, household composition, income, marital status, mother tongue, nationality, ethnicity, religion, sexual orientation, medical identifiers, workplace/ organisation, educational institution, and geographical identifiers. Geographical identifiers include, for instance, postcode, suburb, municipality, province, region, and place where the respondent grew up. (Indirect) identifiers, when triangulated with e.g. geographic locations, IP addresses, postcodes, names of institutions etc. may make it possible to re-identify participants. The greater the number of indirect identifiers held by the researcher, the higher the risk of re-identification. Researchers should therefore minimise data collection and outline how they will mitigate against the risk of re-identification in their research ethics application.

Responses to open-ended questions sometimes contain identifiers which are connected to respondents themselves or other persons, such as name or occupation of a spouse. Disclosure risk must be assessed on a case-to-case basis, with re-coding, pseudonyms or deletion of variables being used if necessary to preserve confidentiality.

The level of anonymisation needed depends on whether a combination of indirect identifiers could lead to the identification of a respondent. If so, then variables can be recoded or deleted to avoid identification: for instance, instead of date of birth, age in months and years could be used; instead of a full postcode, use just the first three characters.

Anonymisation/ pseudonymisation techniques and issues

Changing proper names to codes or pseudonyms is the most popular anonymisation technique used for qualitative data. A good way to keep the anonymisation process under control is to replace personal names with pseudonyms directly after the transcription. Typing a special character in front of all proper names at the initial transcription stage will facilitate the planning and carrying out of anonymisation because all proper names can be easily found within the data.

If retraceable methods, such as key-coding and two-way cryptography are used, the pseudonymised data may still be classified as personal data under the General Data Protection Regulation. The same is true if the researcher, or another person within the University or collaborator, still possesses the means/ key to re-identify participants.[6]

Less well-known anonymisation techniques include swapping and adding random variation to indirect identifiers. Swapping means matching unique cases on the indirect identifier and then exchanging the values of the variable. Please see the UK Data Service's advice on anonymisation and the ICO's guide to data protection.

A diagnosed severe illness can be changed into another, similar type of illness, if doing this does not reduce the usefulness of the data too much. Another option would be to categorise the information in the same way as with quantitative data. For example, 'AIDS' could be changed to [severe long-term illness] and thereafter referred to as [illness], provided that the reader is able to deduce from the context that [illness] refers to the 'severe long-term illness' mentioned at the beginning.

As a general point, in social sciences research, it can be difficult to completely anonymise qualitative personal data without losing its value. Please distinguish between whether you will use full anonymisation or pseudonymisation, both in your CUREC application and the information for participants and consent documents. In most cases, pseudonymisation is the more practical and realistic option, though full anonymisation might be necessary if participants (and researchers) are at risk should participants be re-identified.

[6] See Data Protection & Research guidance, (accessed 19 April 2018)

Examples:

- Identifiable/ personal data: “Mary, 35, 2 children, Brighton”
- Pseudonymised data: “Ruth (i.e. false name), aged 35, 2 children, Brighton”
- Anonymised data: “a parent of two children”

Further anonymisation advice

Detailed logs should be kept of all anonymisation measures carried out. Contact the Research Data Oxford team (via their email address) for more advice about data anonymisation and access control. For further guidance please see ICO guidance on anonymisation and Elliot et al (2016) The Anonymisation Decision-making Framework.

There may be instances where compliance with the GDPR may not be straightforward due to the nature of the research project (e.g. in some Social Anthropology and Ethnography projects, where it may be very easy to re-identify villages/ participants). Issues like these must be addressed in the research ethics application.

Retention of data

Research data and records should be retained for as long as they are of continuing value to the researcher and the wider research community, and as long as specified by research funder, patent law, legislative and other regulatory requirements.

The University policy on the Management of Data Supporting Research Outputs states that:

“Researchers will preserve and provide appropriate access to their research data supporting outputs after the end of their project for as long as it has continuing value, in accordance with legal and funder requirements and paying due regard to discipline norms and cost. Notwithstanding, the minimum retention period for research data and records is three years after publication or public release of the work of the research.”

Note, however, that funders and regulators may require longer retention periods.

In the case of research conducted by students, data retention beyond the duration of their degree course must be discussed and a retention plan agreed with the supervisor.

The GDPR requires that data is not kept as identifiable personal data for longer than is necessary in relation to the purposes for which it is processed. However, personal data processed solely for research purposes, archiving purposes in the public interest, or statistical purposes may be stored indefinitely, provided there are appropriate safeguards in place, such as pseudonymisation. If researchers “justify indefinite retention on this basis, [they] must not later use the data for any other purpose – in particular for any decisions affecting particular individuals.”[7] However, researchers must not hold on to personal data ‘just in case’ this might become useful for the above purposes in future.[8] Research funders and regulators will often have their own specific requirements. In all cases the retention period, or at least its basis and rationale (if not the precise detail), must be communicated to research participants in order to satisfy the GDPR requirement for transparency.

In many instances, researchers will resolve to retain research data and records for a longer period than the minimum requirement. Data archives and institutional repositories (such as ORA-Data at Oxford) are working to address this development. As different regulations apply to how long researchers are required to store records after the completion of research, researchers should look

[7] See ICO’s advice on this (accessed 18 June 2018)

[8] Ibid.

into what repositories might be available to them as a result of their divisional, departmental or institutional affiliations. Researchers must keep research data for the longest applicable period of time or include them as part of a dataset if they are deposited into an archive.

Practical considerations of storage space for data during a project will need to be considered. Expectations and requirements to preserve the data for a long time after the project, when appropriate, will also needed to be planned. This may include instances where researchers wish to reuse their own data for subsequent studies or share it with other researchers after preservation. This situation should be anticipated, and addressed in the original study's information for participants and consent form.

Retention of contact details for future research opportunities

It is common in many research studies for researchers to seek permission to retain contact details of participants in order to offer them the opportunity of taking part in future research. Contacting individuals in the future to invite them to take part in new research projects would be regarded as direct marketing - interpreted broadly as communications that are promotional or advertising in nature and directed at particular individuals.

The rules on direct marketing differ depending on the method by which researchers intend to carry this out. For instance, direct electronic marketing (such as email communications) is subject to the Privacy and Electronic Communications Regulations (PECR) as well as GDPR. Researchers should therefore review the Information Compliance Team's guidance for compliance with the legislation when using mailing lists. This covers considerations such as lawful basis and standard of consent, requirement for transparency via privacy notices and providing a mechanism for individuals to opt-out of receiving such communications.

Management of any mailing list needs to comply with the above legislation, which includes taking a privacy by design approach. In addition to the above points, researchers will therefore need to give consideration to maintaining records of consent, appropriate management of opt-out requests, as well as the general principles of GDPR compliance such as data security, data minimisation, retention and accuracy.

As part of their mailing list databases or registers, researchers may wish to collect and retain other personal attributes in addition to contact details in order to better target communications in the future. If these attributes include special category data such as health/ ethnicity data, then this processing of special category data for the purposes of direct marketing would likely require a Data Protection Impact Assessment (DPIA). Researchers should consult with the Information Compliance Team for advice on the privacy by design process.

Disposal of data

If there are strong reasons why research records need to be destroyed instead of stored and preserved securely, researchers should include additional stages clearly designed to protect participants' confidentiality throughout the process rather than as a set of 'project end' measures. Paper records must be shredded. Records stored on a computer hard drive or USB drive must be erased using commercial software applications designed to remove all data from the storage device. Contact the Research Data Oxford team (via their email address) for more advice about erasing electronic records. For recorded data on CDs, or DVDs or other portable media, the storage devices must be physically destroyed or made un-readable. Local IT support staff periodically hold hard drive destruction 'events', which researchers could take advantage of. Researchers should keep records stating what records were destroyed, and when and how they did so.

Special considerations for audio/ visual data/ photographs

Please first refer to the guidance produced by the Information Compliance Team on remote recording of participants for research projects – see Appendix A.

Increasingly, researchers are in a position to gather data using mixed media that adds new dimensions to the potential for analysis. The value of this needs to be recognised. Where data consist of recordings of individuals, it is especially important to gain explicit consent for audio/ video recording and/ or photography in general, and to gain explicit consent in case the participants are still recognisable (e.g. faces, voices). Audio-visual datasets cannot be easily anonymised. If the datasets contain identifiable information they can be archived only if explicit consent was given for this.

The material recorded may be such that the participant is happy to waive the requirement for confidentiality, and agree that the researcher is free to use the material in any way he/ she chooses, e.g. in public lectures.

Where there is any potential sensitivity of content (e.g. the participant may express views that are private, or demonstrate incompetence in a task), then it is incumbent on the researcher to take extra safeguards. For the majority of projects, points a) and b) below are the most important ones:

a) Informed consent must be in place, which also complies with any data policies of research collaborators (if applicable). The participant information sheet should include that the material will be seen only by members of the research team and other academics (not by members of the public).
b) The relevant recordings must be kept in secure, long-term digital storage, or, for hard copies, in a locked filing cabinet.
c) If depositing sensitive material in an archive, the researcher must work with the archive to ensure that appropriate measures be put in place to restrict access to such material.

In addition, the following safeguards will need to be considered if appropriate:

d) Participants should clarify during recordings any sections that are 'off the record'.
e) Researchers undertake to vet access to data by others (or request that the archive where the data will be deposited undertake this vetting).
f) Special steps will be taken to ensure data is migrated off devices (and fully deleted from them) to secure encrypted storage immediately.
g) When using transcription services, it is important to ensure data is transferred between parties in a secure manner, and that the service deletes all audio-visual material once the transcription has been returned to the researcher. A service-level agreement should be in place before any material is transferred to the service provider. APPEN and Accuro are the University's approved suppliers for the transcription of audio-visual material. The University's Information Security Team has approved the use of the Microsoft Office suite within Nexus365 for processing confidential information, including the automatic transcription feature within Microsoft Teams. This option is considered more secure than many of the third party transcription services on offer.
h) Recordings of children raise two additional ethical issues:
    i. Researchers should be aware that parents and teachers may be concerned that even innocuous recordings of children could be misused, so care must be taken to stress the protections researchers are placing around the data balanced against the benefits of their participation, and the integrity of their research project. Point (b) should be adhered to even when the content of the recording is not apparently sensitive.
    ii. With the passage of time, a child participant may no longer agree to their data being retained. This is unlikely to be a realistic concern except where an adult has given

permission for a video of their child to be made more widely available, e.g. as an illustrative example in a lecture.

i) Researchers should be sensitive to the (rare) possibility of recordings being 'lost' after being archived, and only discovered years later after the researcher who collected the data has disappeared. The researcher should make a plan for the storage and ultimate disposal of the material. Any material that is archived must be labelled as confidential, with the name and contact details of the researcher attached.

For ongoing studies, once child participants have reached an age where they can give their own consent, then this should be sought before making the materials available to those outside the research group.

Resources

Further advice on research data management is available from the Research Data Oxford website, including advice on:

- The University Policy on the Management of Data Supporting Research Outputs
- Working with data, including
  - data management planning
  - data backup, storage and security
- Sharing data
- Tools, services and training

Further advice on data protection from the University of Oxford is available from:

- Data protection & research web pages
- Data protection checklist
- Information Compliance team (via their email address)

Please ensure you have robust research data management plans in place demonstrating a consideration of these points before applying for research ethics review.

Appendix A – Guidance on the Remote recording of participants for research projects

The information below sets out the points of consideration for data protection when remotely recording participants.

Do I need to record?

There is now a demand to be able to hold participant interviews and collect data remotely using video-conferencing tools. However, in the first instance researchers should consider whether there is a need to remotely record participants if it was not necessary to record them prior to a remote working situation. For research, the University generally relies on 'public interest task' as its lawful basis for processing personal data. To rely on this lawful basis, the recording must be necessary for an active research activity and there must be ethical approval in place to conduct that activity.

Video recording in Microsoft Teams

Microsoft Teams is the University's approved tool for virtual meetings and the only tool approved for confidential subject matter. Microsoft Teams has the functionality to video meetings, but researchers will still need to consider the risks when setting up a virtual meeting for the purposes of recording. To find out more, visit IT Services' page on recording meetings.

Necessity and proportionality

Consideration should be given to the necessity and proportionality of the video recording. For example, if a video recording is necessary to capture an assessment of the participant, the video recording should be limited to the assessment only, as it may not be necessary to record the entire meeting for the purpose of the research.

It may only be necessary to record the audio feed of the meeting for the purposes of transcription and later analysis. However, Microsoft Teams does not currently have the functionality to isolate audio from a video recording of a virtual meeting. In order to restrict the recording of the meeting to audio only, all attendees must switch off their cameras before starting the recording. This can only be done by each attendee. The onus is therefore on the participant to disable their own camera feed as it cannot be switched off by the meeting organiser. With this approach, there is a risk that participants may accidentally enable their camera during the recording and the researcher may inadvertently capture their video feed. As a safeguard to ensure that only audio is captured, the invitation email to the invitees could be edited to remind them to ensure their webcams are switched off prior to joining the meeting. Researchers should then remind all attendees in the meeting and check that all cameras are switched off before pressing record.

Where it is necessary to record the audio feed only but researchers need to be able to see the individual during the recording, there will be a risk of over-collection of personal data. Researchers need to consider how to mitigate those risks to avoid processing more personal data than necessary, for example, deleting the video as soon as the transcription is complete, ensuring data security measures are in place to protect the data until it can be destroyed and only using a third party transcription service:

- that has been subject to a third party security assessment (TPSA) and is assessed as low risk for confidential data
- whose contract uses the University's standard template for supply for services or has been approved by the Purchasing Team in accordance with the University's Financial Regulations

Security

Before the meeting starts, researchers should ensure that the working environment is set up appropriately to maintain and protect the privacy and confidentiality of participants, such as using headphones and not allowing unauthorised persons to look over their shoulder.

Once complete, the recordings are saved on Microsoft Stream. The organiser must ensure that the permissions to the recording are set appropriately whilst the recording is stored by Microsoft and restricted to only those with a need to know. For guidance on this, check IT Services' page on recording meetings. The default permissions for the recording are set with the person who made the recording (the meeting organiser) as the owner of the video and, if applicable, the internal Nexus 365 users who were on the meeting invite are set as viewers. External or guest meeting participants will not have access to the recording.

Retention

As the recording will be on an individual organiser's account as opposed to a shared mailbox, it is recommended that recordings are downloaded and saved to the University IT network (for example restricted access folder, password-protected format) for data availability and business continuity purposes and so the retention policy for that data can be easily managed. The recording will exist in Microsoft Stream as long as the owner keeps it there or for as long as their account exists. Once downloaded, the recordings should be deleted from the organiser's individual Microsoft Stream account.

Note that when a user deletes a recording, it is sent to the recycle bin and they have 30 days to recover this before it is permanently deleted. Recordings can also be permanently deleted from the recycle bin before the automatic 30 days.

Transparency

It is important that participants are informed about the proposed recording activities or proposed changes to recording activities for projects already in flight through participant information sheets. It is recommended that the ethics committee are consulted on any changes to participant information sheets.

Alternative to recording in Microsoft Teams

Microsoft Teams virtual meetings can be used to facilitate the interviews with participants, but the audio of the interview be recorded on a separate encrypted dictaphone device (personal mobile phones are unlikely to be appropriate).

Make sure that the working environment is set up appropriately to ensure that all parties can maintain and protect the privacy and confidentiality of data. This is of greater importance if the interview is being recorded over a dictaphone as this will require the audio to be played over the computer speakers. It is recommended that any recordings captured through the device are transferred to the University IT network as soon as possible (for example restricted access folder, password-protected format) and deleted from the device. There are data security risks with this approach, particularly around secure destruction of data held on the device and also the risk of loss of device (and subsequent loss of personal data held on the device) which could result in a confidentiality and an availability personal data breach.

Note that the recording function within Microsoft Teams is switched off by default at the University of Oxford to discourage the inappropriate use of recording.