Central University Research Ethics Committee (CUREC)

Best Practice Guidance 09_Version 3.0

Title: 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 Research Data and Records[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

[1] http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/

[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, including an obligation to obtain the explicit consent of the individual.

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

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 at http://researchdata.ox.ac.uk/. For information about the GDPR and its implications for research please see https://researchsupport.admin.ox.ac.uk/policy/data.

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, how it will be gathered/ transferred/ transcribed, de-identified (if applicable), who will have access to it, where it will be stored and what potential use may be made of the data (e.g. publication, use in future research). Researchers should 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 3 years after publication according to University policy. Please see 'Retention of data' for further information.

For further advice on informed consent, including recommended templates, please see https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/consent

Safe data gathering and storage

While gathering data in the field, mobile devices containing University data should be protected by whole disc encryption. Third party online cloud storage such as Google Drive, Dropbox and OneDrive is **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 researchdata@ox.ac.uk for advice on alternatives.

Personal and special category (formerly known as sensitive) data must be transferred and then stored as safely and securely as possible at the Principal Investigator's University department or faculty, e.g. using encrypted laptops, encrypted USB sticks (for short-term storage), encrypted files in departmental storage systems (e.g. SharePoint) 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 researchdata@ox.ac.uk 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 http://researchdata.ox.ac.uk/home/managing-your-data-at-oxford/storage-and-backup/ 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 is one option (the other is restricting access to data) and should 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) 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.** It is also important not to over-anonymise data. 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 should 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 should therefore be avoided. Terms such as 'fully anonymous' or 'strictly confidential' should be avoided, as they are often impossible to define. Better is to indicate how data will be anonymised (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 https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing and http://www.data-archive.ac.uk/help/user-faq#3/.

Data sharing

Research data may be used or disseminated for research purposes only, and only with the participants' consent. 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 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. Researchers also need to agree with the other party what happens when they no longer need to share the data.[5]

Indirect identifiers / background information

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 years could be used; instead of a full postcode, use just the first 3 digits.

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

[5] See https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/ (accessed 18 June 2018)

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 https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation and https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/.

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.

Further anonymisation advice

Detailed logs should be kept of all anonymisation measures carried out. Contact the Research Data team (researchdata@ox.ac.uk) for more advice about data anonymisation and access control. For further guidance please see http://www.dcc.ac.uk/resources/briefing-papers/legal-watch-papers/sharing-medical-data#4 and http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf.

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 should 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 minimum retention period for research data and records is **three years after publication** or public release of the work of the research according to University policy, though funders and regulators may require longer retention periods.

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 should not hold on to personal data 'just in case' this might become useful for the above purposes in future.[8] Research funders

[6] See Data Protection & Research guidance, https://researchsupport.admin.ox.ac.uk/policy/data/scope#collapse461586 (accessed 19 April 2018)
[7] See ICO's advice on this at https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/ (accessed 18 June 2018)
[8] Ibid.

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), should be communicated to research participants in order to satisfy 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 into what repositories might be available to them as a result of their divisional, departmental or institutional affiliations. Researchers must keep records 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.

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 their participants' confidentiality throughout the process rather than as a set of 'project end' measures. Paper records should be shredded. Records stored on a computer hard drive should be erased using commercial software applications designed to remove all data from the storage device. Contact the Research Data team (researchdata@ox.ac.uk) for more advice about erasing electronic records. For data stored on USB drives or recorded data on CDs, or DVDs or other portable media, the storage devices should 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.[9]

Special considerations for audio / visual data / photographs

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

[9] Based on advice from http://www.virginia.edu/vpr/irb/sbs/resources_guide_data_retention.html

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 should be kept in secure, long-term digital storage, or, for hard copies, in a locked filing cabinet.

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

c) Participants should clarify during recordings any sections that are 'off the record'.

d) Researchers undertake to vet access to data by others.

e) 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.

f) Special steps will be taken to ensure data is migrated off devices (and fully deleted from them) to secure encrypted storage immediately.

g) 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 should 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.
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 Oxford Research Data website at http://researchdata.ox.ac.uk/, including advice on:

- The University Policy on research data management
- 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
- The data protection team: data.protection@admin.ox.ac.uk

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