Specialist Research Ethics Guidance Paper

PRINCIPLES OF ANONYMITY, CONFIDENTIALITY AND DATA PROTECTION

Note: This guidance document aims to develop further the information relating to anonymity, confidentiality and data protection that is covered in the University’s ‘Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue’, in line with changes to data protection legislation during 2018 (the implementation of the General Data Protection Regulation (GDPR) and the UK Data Protection Bill 2018), and provides signposting to guidance from useful external sources, in particular the Information Commissioner’s Office, the Health Research Authority (HRA), which has produced useful guidance on implementing the GDPR for health research, the Medical Research Council’s Regulatory Support Centre, and the EU’s Article 29 Working Party. A useful practical summary by the HRA on the implications of the GDPR for research in the UK can be found here.

This document will be updated as further changes to legislation are put in place, and more guidance becomes available.

In summary:
If you are processing (i.e. collecting, storing, using, disclosing or destroying) identifiable personal information about living individuals, then you should ensure that you comply with the requirements of the General Data Protection Regulation (GDPR) and the Common Law Duty of Confidentiality. In the UK, once the Data Protection Bill becomes law, its requirements will also need to be met (staff and students working at the International Faculty in Greece will need to ensure that any relevant local data protection regulation is met in addition to the GDPR).

If you are processing (i.e. collecting, storing, using, disclosing or destroying) identifiable personal information about deceased individuals, then you should ensure that you comply with the requirements of the Common Law Duty of Confidentiality. You should also be aware of the possibility of living individuals (e.g. relatives of the deceased) being identified in this information, which would then need to be treated in line with the relevant data protection legislation as stated in the previous paragraph.

If you are processing (i.e. collecting, storing, using, disclosing or destroying) anonymised personal information, whether relating to the living or the deceased, then your research activity falls outside the scope of these guidelines.

The use of identifiable personal information in research should be reduced so far as possible. You should think carefully about how it may be possible to use less identifiable data (e.g. rather than collecting full date of birth, would it be sufficient to collect only ‘month and year’? Is it necessary to collect, or retain, the full post-code?). All processing of personal information should be defensible as both relevant and accurate.

If it is necessary to use identifiable personal information, you should aim at all times to ensure that the processing is defensible as both ‘fair, lawful and transparent’. This requires you to be as transparent as possible about the uses to which data will be put and any risks involved. The data subject (i.e., the individual whose data are being processed) should be fully informed about how and why their data will be processed, including the legal basis for the processing (for most research this will be ‘a task in the public interest’; additional conditions apply to Special Categories of personal data). You should usually only use identifiable personal information with the consent of the data subject. It may be possible to use such data without consent, providing consent is not being used as the legal basis for the processing (e.g. in the
case of research involving large datasets obtained from social media, where it may be infeasible to seek informed consent from all individuals concerned); however, consent is to be preferred unless it can be shown to be inappropriate for some reason.

You should ensure that personal information is kept secure at all times. The level of security should be proportionate to the risks but all personal information should be kept securely.

You should not keep personal information for longer than necessary; however, it is recognised that (as long as relevant conditions are satisfied) research may require the retention of data for long periods and this may be justified (e.g. to meet legal or funder requirements).

You should avoid disclosing identifiable information, including information that may be identifiable to others, wherever possible. If it is necessary to disclose personally identifiable information, or information that may be potentially identifiable, then this should usually only be done with the consent of the individual/s involved.

1. Introduction

A researcher who processes (i.e. collects, stores, uses, discloses or destroys) identifiable personal information - as defined in the box below - about living individuals, must comply with the requirements of the relevant data protection legislation, and the Common Law Duty of Confidentiality.

A researcher who processes identifiable personal information about deceased individuals, must still consider the requirements of the Common Law Duty of Confidentiality. Individuals have a reasonable expectation of privacy with respect to confidential information that refers to them. Any use of such confidential information that exceeds that which an ordinary person could reasonably be said to expect constitutes a breach of confidentiality. In addition, researchers should be aware of the possibility of living individuals (e.g. relatives of the deceased) being identified in this information, which would then need to be treated in line with the relevant data protection legislation as stated in the previous paragraph.

From 25 May 2018, the relevant data protection legislation in the EU (including the UK) will be the GDPR, and it is expected that the requirements of the GDPR will continue to apply in the UK after it leaves the EU. In addition, a new UK Data Protection Bill is progressing through Parliament which will make some specific changes/additions to the GRPR requirements, as they apply in the UK. Staff and students working at the International Faculty in Greece will need to ensure that any relevant local data protection regulation is met in addition to the GDPR.

The new legislation strengthens the rights of the individual whose data are being processed (the ‘data subject’), but also incorporates a range of exemptions from these rights for research purposes, providing appropriate safeguards are in place (e.g. there will normally be no right for research participants to access their data, rectify it or have the data erased, if this would prevent or seriously impair the achievement of the research purpose). For more guidance, see the Health Research Authority’s guidance on ‘Data Subject Rights and Research Exemptions’.

Any processing of personal data must have a defined ‘Data Controller’ in place (the individual or organisation which determines the purposes and means of processing personal data). For research undertaken by staff or students of the University of Sheffield, the Data Controller will usually be the University of Sheffield (i.e. not a particular individual or research team). Collaboration with other institutions may result in alternative or joint Data Controllers; there
should be agreement of which organisation(s) take on this responsibility at the outset of a research project, and this should be clearly documented via collaboration agreements.

Data protection legislation applies to ‘personal data’. This is defined in the General Data Protection Regulation (GDPR) as:

> ‘any information relating to an identified or identifiable natural (living) person (‘data subject’): an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.’

The processing of fully anonymised personal information, whether relating to the living or the deceased, falls outside the scope of these legal requirements. Fully anonymised data are those from which the original data subject cannot be identified by any member of the research team, using either the dataset itself, or any other dataset that may be accessed by members of the research team.

In practice, in the case of discrete research projects, it is highly unlikely that members of the research team will come into contact with data from other parts of the University that may result in the re-identification of participants whose data have been anonymised. However, researchers should think carefully about this possibility when seeking to anonymise their data; strictly speaking, if there is any possibility that anonymised data could be traced back to the data subject via any other data held by, or likely to come into the possession of, the Data Controller, then the data has in fact only been ‘pseudonymised’. This means that it would in fact still be classed as personal data. Two examples of situations in which this problem is more likely to arise include:

- administrative research, in which research staff may have access to central University records that may link data to the participants that provided it;
- types of research in which there are particular identifiers that are widely used outside the research team (e.g. health research involving NHS numbers).

In addition, you should also be aware that, if the research team encompasses individuals from other organisations as part of a research collaboration, those individuals may have access to datasets that may enable the identification of participants.

The use of identifiable personal information in research should be reduced as far as possible, consistent with achieving the research aims. Thus researchers should always think carefully about (a) whether it is necessary to use identifiable personal information, (b) what is the earliest stage at which de-identification might be possible without compromising the integrity of the research and (c) how full, robust anonymisation can be achieved. All uses of personal information should be defensible as accurate, relevant and not excessive.

For more guidance on anonymisation, refer to the ICO’s ‘Anonymisation Code of Practice’ (this is due to be updated soon, in line with GDPR but is still a helpful document).

2. Identifying an appropriate legal basis for the processing of personal data

If it is necessary to use identifiable personal data, then an appropriate legal basis for the processing of this data must be identified, and researchers must be explicit about this and
document it as part of their ethics application, and in the information they provide to participants.

Article 6 of the GDPR sets out six possible legal bases for processing of data that does not include ‘Special Categories’ (these are discussed later in this document and have additional requirements). At least one of these legal bases must apply whenever personal data is collected and used as part of a research project.

The University’s view is that, for the vast majority of research undertaken at the University, the appropriate legal basis will be:

**6(e) Public interest: the processing is necessary for you to perform a task in the public interest or for your official functions.**

Further details are set out in the University’s Privacy Notice, and a link to this can be included in the information that is provided to participants: https://www.sheffield.ac.uk/govern/data-protection/privacy/general. Other legal bases are available and may apply to other aspects of University business, but are unlikely to apply for research purposes. If you feel that the research you are undertaking cannot be justified as being ‘a task in the public interest’, please contact the University Research Ethics Committee for further guidance.

Although the legal basis for processing a person’s data is most likely to be ‘a task in the public interest’ rather than ‘consent’, from an ethical perspective, obtaining a person’s informed consent for their involvement in the research is still likely to be required, unless it can be shown to be inappropriate for some reason (e.g. if the material is already in the public domain, for example). If a researcher intends to process data without informed consent, then further advice should be sought from the University Research Ethics Committee.

Further guidance on legal bases is provided in the HRA’s guidance document: *A Lawful basis for health research under the Data Protection Law* and the MRC’s guidance document: *Guidance note 3: GDPR, Consent in Research and Confidentiality*.

### 3. Data Protection Safeguards

’Safeguards’ are measures to protect the rights and freedoms of individuals whose personal data you are processing. Under the GDPR there is greater emphasis on implementing safeguards for research. In practical terms, this means giving careful consideration to:

- Only collecting personal data where it is necessary for the research purpose (known as ‘data minimisation’);
- Ensuring that data are pseudonymised or anonymised wherever possible and as early as possible;
- Ensuring appropriate arrangements are in place for security and storage of data, proportionate to the risks inherent in the nature of the data e.g. portable devices must be encrypted.

For processing of ‘Special Category’ personal data, additional safeguards will be required: Further information about this can be found in section 6 of this document: ‘Research involving ‘Special Categories’ of personal data’.

It should be noted that safeguards will not be sufficient if the processing is likely to cause substantial damage or distress to an individual. In addition, currently, the GDPR states that
safeguards will not be sufficient if the processing is ‘carried out for the purpose of measures or decisions with respect to a particular data subject’. An amendment stating that this will be allowed if the processing is for the purpose of ‘approved medical research’ (i.e. approved via the HRA, NHS research ethics committee, etc.) is being proposed as part of the Data Protection Bill, to prevent a negative impact on interventional health research.

More guidance on safeguards can be found in the Health Research Authority’s guidance document: ‘Data Protection Safeguards’.

4. The right to be informed

When gathering identifiable personal data researchers should aim at all times to ensure that its processing is defensible as ‘fair, lawful and undertaken in a transparent manner’. This requires that the participant be provided with appropriate information about the uses to which data will be put and any risks that might be involved. This information must be:

- Concise, transparent, intelligible
- Provided in easily accessible form, using clear, plain language
- Prepared in consideration of the needs of the audience e.g. information addressed specifically to a child
- Provided by an appropriate means (e.g. in writing, electronically, orally)

Under the GDPR, this information should specifically cover the legal basis that is being applied in order to process someone’s personal data. In many contexts, taking into account the language and literacy of potential participants, a fact-sheet (often referred to as a participant information sheet) is a useful and documented means of providing this information. However, a ‘layered’ approach to providing this information may be useful (e.g. utilising webpages, posters, leaflets or newsletters as well as information sheets).

Taken together, the information provided should normally include:

- the nature and purpose of the project;
- the legal basis for the collection and use of the participants’ data (and the additional condition(s) required for processing of ‘Special Categories’ of data, if required);
- the research methods to be employed by the project;
- full explanation of any technical terms used;
- the conditions under which the project will be conducted;
- who is undertaking and who is sponsoring the project (i.e. the details of the ‘Data Controller’, the research team, the funder and/or the research governance sponsor if applicable);
- the potential risks and inconveniences that may arise;
- the potential benefits that may result;
- what participation in the research will require in practice and what data will be collected;
- information about the right to withdraw from the research, and how to go about this;
- what will happen to the data and who will have access to it (including any further use of the data beyond the immediate research project, any intention to transfer data to any third party, and the appropriate safeguards which will be adopted);
- how participant confidentiality will be safeguarded;
- how the data will be stored, and when it will be destroyed (or the criteria that will be used to determine when it will be destroyed);
• how to raise concerns, or to complain, about the research, and to whom (see note below); and
• the consequences of non-participation (such as alternative treatments in the case of some medical research, or alternative educational activities in the case of some educational research).

In connection with the above, it should be noted that the appropriate channels for the registration of complaints within the University, should a participant be unhappy with their treatment and unable to resolve them directly with the researcher and/or research team, is the Head of the relevant department. Participants should also be informed of their right to contact the Data Protection Officer for the Data Controller organisation, or the Information Commissioner’s Office, if they have a complaint about the use of their personal information within the research.

Personal information should not be retained for longer than necessary. However, it is recognised that research may require the retention of data for long periods and that this may be justified, for example due to funder requirements. The participant should be given full information about how their data will be used, how it will be stored and for how long (if the latter is not possible, then the participant should be informed of the criteria that will be used to determine retention periods.)

5. Research involving ‘Special Categories’ of personal data

Sensitive data is referred to as ‘Special Category’ data in the GDPR and UK Data Protection Bill, and includes:

• data revealing racial or ethnic origin,
• political opinions, religious or philosophical beliefs, or
• trade union membership;
• the processing of genetic data or biometric data for the purpose of uniquely identifying a natural (living) person;
• data concerning health;
• data concerning a natural person’s sex life or sexual orientation; or
• criminal records or allegations of criminal/illegal activities

[It should be noted that from an ethical perspective, the University considers a number of other types of data to be sensitive: the full list can be found in Research Ethics Policy Note no.6 ‘Research involving vulnerable people’:  

In order to process Special Category data lawfully, you must identify both a lawful basis under Article 6 of the GDPR (set out in section 2 of this document) and a separate condition for processing Special Category data under Article 9(2). The condition most likely to apply to research is:

9(2)(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

In order for this condition to be relied upon, it must be justifiable that this processing is in the public interest, must not cause substantial damage or distress to the data subject, and appropriate safeguards must be in place (e.g. processes to ensure data security - see section 3 of this document for details of safeguards). The University’s view is that the information provided in an ethics application (e.g. concerning the aims and objectives of the research), and
the assessment of this via the process of ethical review, will meet the researcher’s obligations in respect of the need to justify that the research is in the public interest.

Other conditions which may apply in certain circumstances (researchers should contact the University Research Ethics Committee for advice if they wish to rely on these) are:

9(2)(a) the data subject has given explicit consent to the processing of the personal data for one or more specified purposes (ONLY TO BE USED IF NO OTHER CONDITION IS POSSIBLE – MORE STRINGENT CONSENT REQUIREMENTS WILL APPLY);

9(2)(e) processing relates to personal data which are manifestly made public by the data subject (this may apply when using certain social media data, for example);

9(2)(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices.

Researchers who need to process Special Category personal data as part of their projects must explicitly state which legal basis AND which condition they are relying on, as part of their ethics application, and in the information they supply to participants.

Further guidance on the use of Special Categories of personal data for research purposes can be found in the Health Research Authority’s document ‘A Lawful Basis for Health Research under Data Protection Law’. This is primarily intended for health research but has wider applicability.

6. **Re-using personal data for a different purpose & sharing with third parties**

If a researcher wishes to re-use personal data that were collected for a particular purpose (e.g. a specific research project) for a new purpose (e.g. a new research project), and the data subject was not informed of this as part of the original informed consent procedures, then the researcher would be required to contact the data subject to inform them of this BEFORE the new processing commenced. If the data from the original project had already been fully anonymised before use in the second project, it would no longer constitute personal data and would therefore no longer be subject to data protection legislation and the data subject would not need to be contacted about the re-use of their data.

Where personal data is to be used by a researcher but they have NOT obtained the data directly from the data subject, the original data controller supplying the data must have informed the data subject of relevant information relating to this new processing. However, the receiving data controller should check that the providing data controller has met their obligations in this regard, and it is also good practice for the receiving data controller to provide relevant study-level information to the data subject.

Guidance on the relevant information that should be provided to the data subject in these circumstances, and the appropriate time frames for providing this information, are provided in the Health Research Authority’s guidance document: ‘Transparency, Health Research and Data Protection Law’.
In some circumstances, where personal data has NOT been obtained directly from the data subject, then the requirement to provide information to the data subject does not apply. This is where:

- Data has been pseudonymised and the new research activity is conducted without using identifiable data AND
- The provision of information would be impossible or involve a disproportionate effort (taking into consideration the number of participants, the age of the data, etc.) OR
- The provision of information would render impossible or seriously impair the objectives of the research.

Such a decision should be documented as part of the ethics review procedure. Where information is not provided to the data subjects due to the above exemptions, the information should instead be made publicly available (e.g. via a study webpage).

7. **Common Law Duty of Confidentiality**

The Common Law Duty of Confidentiality applies to research involving confidential personal information. Under the law of confidentiality, it is recognised that individuals have a reasonable expectation of privacy in relation to confidential information: any use of confidential information that exceeds that which an ordinary person could reasonably be said to expect will constitute a breach of confidence.

Information will be considered confidential if an individual could be understood to have an objective reasonable expectation that the information will, in the circumstances, be kept private.

The easiest way to affect an individual’s reasonable expectations is by explaining clearly what will happen with their personal information. Minimally, it should be made clear who will have access to their data, for what purpose(s), and for how long. Special considerations apply, and further specific advice should be sought, if considering seeking consent from children (0-18), from vulnerable persons with capacity to consent, and vulnerable persons without capacity to consent. Further information about these issues can be found in Research Ethics Policy Note no.2 ‘Principles of Consent’:

[https://www.sheffield.ac.uk/polopoly_fs/1.112749!/file/Research-Ethics-Policy-Note-2.pdf](https://www.sheffield.ac.uk/polopoly_fs/1.112749!/file/Research-Ethics-Policy-Note-2.pdf)

If the intention is to use confidential information for a research purpose, then that should be clearly explained to an individual and their consent, either express or implied, sought for such use.

It should also be made clear to an individual that, wherever possible, there is an ongoing entitlement to withdraw consent to the processing of data for specific purposes. It may not always be possible to grant participants the entitlement to withdraw – for example if data have been anonymised, once publication has taken place, if participatory research processes mean that withdrawal will invalidate the contributions of others. The parameters within which withdrawal is possible should be explained to participants.

Researchers must make clear to participants any intention to provide third party access to confidential information, including after the project’s conclusion. Without an express indication to the contrary, access must be restricted to the lead researcher and researchers directly involved in the research.
A researcher may not disclose the identity of a person, or disclose any information that could identify that person, without having obtained, in advance, that person’s consent to do so, preferably in writing. If the research process is such that it is unavoidable that participants may in some circumstances be explicitly identified to others, then the researchers should explain why this is the case, and described what precautions may be taken (for example, the case of a focus group, this may include discussing confidentiality with all participants at the start of the session and asking individuals not to report details of what has been discussed outside of the group).

Researchers should be aware of the risks to anonymity, confidentiality and privacy posed by technologies of personal information storage and processing which directly identify a person: audio and video recordings, electronic and paper-based files, and e-mail records. Measures to prevent accidental breaches of confidentiality should be taken. Provisions for data security at the end of a project must be made (see the University of Sheffield’s Good Research & Innovation Practices policy, Sections 3.1: https://www.sheffield.ac.uk/rs/ethicsandintegrity/index).

The use of confidential information in ways that are consistent with a valid consent will not represent a breach of confidentiality. However, it may be that there is no valid consent to rely upon and re-contact for fresh consent is not practicable. Any researcher considering this is advised to contact the UREC for further advice.