DATA TYPE AND IDENTIFIABILITY

The identifiability of research data is an important consideration for any research project involving human participants. An understanding of data and identifiability is necessary to manage the risks involved in collecting, handling, and storing research data pertaining to participants.

This Supplementary Guidance is designed to help you consider the type and identifiability of data collected in your research so that you can consider and manage or mitigate the risk of harm or discomfort to research participants or others who may be at risk in accordance with the ethical principles contained in the [National Statement on the Ethical Conduct of Human Research 2023](#) [(National Statement)](#).

What is data?

For the purposes of the National Statement, '**data**' is intended to refer to bits of information in their raw form, whereas '**information**' generally refers to data that have been interpreted, analysed, or contextualised.[1] (emphasis added)

The National Statement recognises that human research uses multiple methods to create, collect, use and store to achieve research goals. Where that data can have an impact on the research participant, the ethics review body wants you to consider this in advance.

While there is no agreed definition of, or right to, 'privacy' in an Australian context there are laws which speak to the 'privacy' of certain types of data or information, and which also regulate the ways in which it may be collected, used, shared, stored, re-used, and destroyed[2].In a human research context, the following types of data or information are protected:

- **Personal information:** Any information or opinion recorded in any form and whether true or not, about an identifiable individual or from which their identity can be reasonably ascertained (for example, name, address, mobile phone number, email address, photo, voice recording, employment record, student record, medical record etc.).
- **Sensitive information:** A special category of personal information that might be used to discriminate against an individual and therefore requiring more protection (for example, racial or ethnic origin, sexual preferences or practices, political opinions, membership of a political association, religious beliefs or associations, philosophical beliefs, union membership or criminal record).
- **Health information**: Defined broadly as information or opinion about, physical, mental, or psychological health, disability, expressed wishes about health care provision, a health care service provided, of an identifiable individual (living or dead).
- **Other:** Data when, on its own, does not meet the definitions above and does not identify a participant but when linked with other data makes identification by others possible.

What do we mean by the identifiability of participants data?

The identifiability of participants in research data is essential for any research project. Understanding data and identifiability is necessary to manage the risks involved in collecting, handling, and storing research data about participants. Personal, health and sensitive information should be considered subject to relevant privacy legislation[3] Any personal information about participants collected in a research project must comply with relevant privacy legislation. The Privacy and Data Protection Act 2014 (Vic) defines personal information as:

---

[1] [National Statement on the Ethical Conduct of Human Research 2023](#)
[2] [Privacy Act 1988](#), [Victoria Privacy and Data Protection Act 2014](#)
[3] [Privacy Act 1988](#), [Victoria Privacy and Data Protection Act 2014](#)

> "information or an opinion (*including information or an opinion forming part of a database*), that is recorded in any form and whether true or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion, but does not include information of a kind to which the Health Records Act 2001 applies."

While the National Statement does not use the terms 'identifiable', 're-identifiable' or 'non-identifiable' as descriptive research data categories due to ambiguities in their meanings, you will see these in research language and in the Research Ethics Platform application form. The terms are more applicable to the process used in the management of your data.

Where the terms identifiable, re-identifiable and non-identifiable are used in relation to kind of data being collected, used, or stored the meanings are as follows:

– **Identifiable**: Identity of an individual can be reasonably ascertained.

If your research project involves gathering participants' personal information – such as names and contact information – then you are working with identifiable data even if that data will not appear in research outputs. Common research methods such as interviews, for example, almost always involve collecting identifiable data because interviewees are asked to record their name and signature on consent forms. Focus groups always involve identifiable data not only because participants almost always sign a consent form, but also because the nature of a group discussion is such that participants will be aware of who said what.

– **Re-identifiable**: Possible to re-identify an individual. For example, identifiers removed from main dataset and replaced by Code (stored separately) which enables identification.

Re-identifiable data is data which enables the identity of individual participant(s) to be ascertained through an indirect means. Most commonly this involves a dataset in which identifiers have been removed and replaced by a code, for example replacing participants' names with a number. The key to the code is stored separately; individual participants can only be identified by accessing both the dataset and the separately stored code.

– **Non-identifiable**: Any identifiers permanently removed, and no specific individual can be identified.

An anonymous survey, for instance, involves non-identifiable data because no identifying information is collected. Note, if an option is given for participants to provide their contact details for any reason (such as a follow-up interview), in that case, identifiable data may be collected.

What are some important ethical considerations around participant data?

In recruitment: When working with identifiable or re-identifiable data, it is important to carefully consider the way in which personal information will be collected. For instance, the recruitment method known as 'snowballing' – whereby existing research participants refer their contacts as potential participants – can contravene privacy laws. This is because the researcher cannot be certain that those contacted by existing participants have consented to their contact information being shared. An ethical means of snowballing involves having participants pass on information about the research to their contacts; if those people then wish to participate, they may contact the researcher directly.

In the de-identification process: When considering this process, you must ensure that all relevant information is coded, that the key is securely stored separately from the data storage, that access

permissions are documented and controlled. This should be outlined as a part of your ethics application and included in your data management plan.

What are some important considerations with maintaining research approvals?

Non-participant Data: Often we do not set out to collect identifiable data as part of our research, yet we may become aware of it through the course of data collection. For instance, participants can divulge personal information about other people (non-participants) that may allow the identity of those people to be ascertained by others. Where this is the case, steps may need to be taken to minimise the risk of the identification of these non-participants. This might involve things such as redacting transcripts or obscuring identifying details in research outputs.

Storage: Data storage, for the lifecycle of the research project and retention period in accordance with university policy and relevant privacy legislation, should be considered and privacy considerations adhered to and documented in the research data management plan and ethics submission. For guidance on how to handle and store data and information, see the La Trobe University Research Data Management Procedure. Knowledge of appropriate data management techniques is required to submit an ethics application.

Key things to remember when completing your application:

What type of data will you be collecting from participants and does this change from what you will be using in the research?

Reflecting on the types of information listed in this guidance note and the examples given. Will you be collecting names, addresses or phone numbers? Then you are collecting *personal information*. If you are collecting physical or mental health diagnoses than you are collecting *health information*.

Sometimes you may collect personal information, but you do not use it in your research – such as names and email addresses. You may use these for scheduling interviews where you collect other data, but you do not use the personal information, the ethics committee needs to know this.

What is the **identifiability** of the data you will be **collecting** from participants and does this change from what you will be **using** in the research?

Over the lifecycle of a research project or activity the identifiability of data may change. For example, it may be collected in an identifiable format, made re-identifiable (coded) for the purposes of use and analysis and then stored in a non-identifiable format (separate file containing codes destroyed and all identifiers permanently removed).

Further information

For further advice on this topic or other human research ethics matters, please email humanethics@latrobe.edu.au. An Ethics Advisor will assist you and may connect you to one of the LEAP or HREC members in your discipline who can offer expert ethics advice.