

Ethics, consent & data sharing

SHARE PERSONAL AND CONFIDENTIAL RESEARCH DATA ETHICALLY

Sharing of research data that relates to people can often be achieved using a combination of obtaining consent, anonymising data and regulating data access.

This guide is intended for those engaged in research involving human subjects which is subject to ethics approval, and for those with a role to play in the Human Research Ethics Committees which oversee the research.

Research institutions provide extensive guidance to researchers about ethical issues and requirements. This does not always include information about the relationship between ethical research and data sharing, and how data can be shared ethically and legally. This guide is intended to go some way to filling that gap.

Contents

1. Introduction	1
2. Meeting your obligations.....	2
3. Informed consent and data sharing.....	3
Case Study - The Australian Data Archive (ADA)	4
4. Human Research Ethics Committees and data sharing	5
Case Study - Victoria University: A journey to an ethics-friendly data-sharing policy	6
5. Anonymising data	7
Audio-visual data	8
6. Access control and licensing	8
Case Study - The Aboriginal and Torres Strait Islander Data Archive (ATSIDA).....	9
Licensing.....	10

1. Introduction

When research involves obtaining data from people, researchers are expected to maintain the high ethical standards set out in the *National Statement on Ethical Conduct in Human Research*¹, both during research and when sharing data.

Research data—even sensitive and confidential data—can be shared ethically and legally if researchers pay attention, from the beginning of research, to four important aspects:

- including provision for data sharing when gaining informed consent
- protecting people’s identities by anonymising data where needed

¹ <http://www.nhmrc.gov.au/guidelines/publications/e72>



- considering controlling access to data
- applying an appropriate licence

These measures should be considered jointly. The same measures form part of good research practice and data management, even if data sharing is not envisioned. Data collected from and about people may hold personal, sensitive or confidential information. This does not mean that all data obtained by research with participants is personal or confidential.

Why share your data? There are many good reasons to share data and benefit you, your discipline (and possibly others) and the research institution in which you work. Sharing your data allows:

- New discoveries from existing data
- Integration of sets of data for new analysis
- Re-analysis of expensive, rare or unrepeatable investigations
- A DOI (Digital Object Identifier) to be assigned to your data so that it can be cited, its use tracked in the same way as journal articles and your research recognised and rewarded

2. Meeting your obligations

In Australia, research on human subjects is covered by the *National Statement on Ethical Conduct in Human Research* which recognises the value of making data available for future research.

Strategies for dealing with confidentiality depend upon the nature of the research, but are essentially informed by a researcher's ethical and legal obligations. A duty of confidentiality towards informants may be explicit, but need not be.

Legislation that may impact on the sharing of confidential data:

- Privacy Act 1988 (Cth) (and state equivalents)
- Human Rights Act 2004 (Cth) (and state equivalents)
- Freedom of Information Act 1982 (Cth) and amendments in the Freedom of Information Amendment (Reform) Act 2010 (Cth) (and state FOI and Right to Information (RTI) equivalents)

DEFINITIONS

Personal data

Personal information is information that identifies or could identify someone. There are some obvious examples of personal information, such as name or address. Personal information can also include medical records, bank account details, photos, videos, and information about personal preferences, opinions and occupation - basically, any information where someone may be reasonably identifiable.*

The *Privacy Act 1988* (Cth) defines personal information as:

*"... information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion." ***

Confidential data

Confidential data is data given in confidence or data agreed to be kept confidential (i.e. secret, between two parties) that is not in the public domain such as information on business, income, health, medical details, and political opinion.

* Based on information provided by the Australian Privacy Commissioner at <http://www.privacy.gov.au/aboutprivacy/what>

** Quoted at <http://www.privacy.gov.au/aboutprivacy/what>



Researchers are also bound to meet the guidelines of the *Australian Code for the Responsible Conduct of Research*² and obligations set out in funding rules.

The *Australian Code for the Responsible Conduct of Research* states that:

- The potential value of the material for further research should [...] be considered, particularly where the research would be difficult or impossible to repeat (Section 2)
- Research data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters (Section 2.5.2)
- Researchers have a responsibility to their colleagues and the wider community to disseminate a full account of their research as broadly as possible (Section 4.4)

The *Australian Research Council (ARC) Funding Rules* for 2012 state:

The Final Report must justify why any publications from a Project have not been deposited in appropriate repositories within 12 months of publication. The Final Report must outline how data arising from the Project has been made publicly accessible where appropriate.³

The revised Funding Agreement of the National Health and Medical Research Council states:

If required by an NHMRC policy about the dissemination of research findings, the Administering Institution must deposit any publication resulting from a Research Activity, and its related data, in an appropriate subject and/or open access repository (such as the Australian Consortium for Social and Political Research Inc. archive or databases listed under the National Centre for Biotechnology Information) in accordance with the timeframe and other requirements set out in that policy.⁴ (Paragraph 12.9)

3. Informed consent and data sharing

Researchers are expected to obtain informed consent for people to participate in research and for use of the information collected. Wherever possible, the value of the data to the wider research community should be taken into account during the planning process and consent forms designed accordingly, as should data preservation and longer-term use. At a minimum, consent forms should not preclude data sharing, such as by promising to destroy data unnecessarily. Researchers should:

- inform participants how research data will be stored, preserved and used in the long-term
- inform participants how confidentiality will be maintained, e.g. by anonymising data
- obtain informed consent, either written or verbal, for data sharing

To ensure that consent is informed, consent must be freely given with sufficient information provided on all aspects of participation and data use. There must be active communication between the parties. Consent must never be inferred from a non-response to a communication such as a letter. Without consent for data sharing, opportunities for sharing research data with other researchers can be jeopardised. While consent can in most cases be obtained retrospectively, this is time-consuming and inconvenient.

The *National Statement*⁵ defines three levels of consent for the future use of data which must be made

² <http://www.nhmrc.gov.au/files/nhmrc/publications/attachments/r39.pdf>

³ http://www.arc.gov.au/pdf/LP12/lp_funding_rules_2012_19%20April%202011.pdf, Section (13.3.2).

⁴ http://www.nhmrc.gov.au/files/nhmrc/file/grants/admin/nhmrc_funding_agreement_110812%20.pdf

⁵ <http://www.nhmrc.gov.au/files/nhmrc/publications/attachments/e72.pdf>, p21.



clear to the research subject:

- 'specific': limited to the specific project under consideration
- 'extended': given for the use of data or tissue in future research projects that are either (i) an extension of, or closely related to, the original project; or (ii) in the same general area of research (for example, genealogical, ethnographical, epidemiological, or chronic illness research);
- 'unspecified': given for the use of data or tissue in any future research.

In addition, the consent form must specify whether the data is to be held in a form which is identifiable, non-identifiable or re-identifiable. Keep in mind that some data, such as human tissue samples, is, at least in principle, re-identifiable (our genetic makeup is unique) so you may wish to provide an assurance that access to the data will be strictly controlled.

Your local ethics office will be able to provide you with more information about written and verbal consent and other matters to consider in constructing consent forms.

Case Study - The Australian Data Archive

The Australian Data Archive (ADA) holds data from the social sciences.

Survey and qualitative data held by the Australian Data Archive (ADA) is always anonymised, unless specific consent has been given for personal information to be included. Metadata about the research data (including study investigators, methodology and access conditions) is made available to the public, but the raw data is not in the public domain and use is regulated for specific purposes after user registration. To access data, users must register with ADA and complete an access request application. Once approved, they must also sign a user undertaking, agreeing to conditions such as not attempting to identify any individuals from the data (through, for example, data mining) and not sharing data with unregistered users. For confidential or sensitive data, stricter access regulations may also be imposed.

www.ada.edu.au

ADA Qualitative aims to facilitate data sharing among qualitative researchers under rigorous access and use policies determined by the depositor. Data can be deposited in virtually any format, including document, audio and video files. The integration of qualitative data into the ADA network more broadly also means researchers using mixed methods combining quantitative and qualitative data will be able to keep their data together in the one archive. We also provide guidelines for data users in areas such as citation of archived datasets in published papers and designing a project methodology based on the use of secondary data.

[**www.ada.edu.au/qualitative/home**](http://www.ada.edu.au/qualitative/home)



DEFINITIONS

Individually Identifiable Data

Data from which the identity of a specific individual can reasonably be ascertained.

Non-identifiable data

Data that have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of non-identifiable data are those that can be linked with other data so it can be known they are about the same data subject, although the person's identity remains unknown,

Re-identifiable Data

Data from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets.

Quoted from definitions in the *National Statement on Ethical Conduct in Human Research*
<http://www.nhmrc.gov.au/files/nhmrc/publications/attachments/e72.pdf>, pp100-101

4. Human Research Ethics Committees and data sharing

The role of Human Research Ethics Committees (HRECs) is to help protect the safety, rights and well-being of research participants, and to promote ethically sound research. This involves ensuring that research complies with legislation regarding the use of personal information collected in research. In research with people, there can be a perceived tension between data sharing and data protection where research data contain personal, sensitive or confidential information. However, in many cases, data obtained from people can be shared while upholding both the letter and the spirit of data protection and research ethics principles. HRECs can play a role in this by advising researchers that:

- most research data obtained from participants can be successfully shared without breaching confidentiality
- it is important to distinguish between personal data collected and research data in general
- privacy laws do not normally relate to anonymised data
- personal data should not be disclosed, unless consent has been given for disclosure
- identifiable information may be excluded from data sharing
- many funders, such as the Australian Research Council and the National Health and Medical Research Council, require that data resulting from funded research should be made available for future research
- even personal sensitive data can be shared if consent has been obtained and if suitable procedures, precautions and safeguards are followed
- there are data archives available which can provide appropriate access controls and secure data storage
- good data management involves careful data management planning.

HRECs can play a critical role by providing such information to researchers, at the consent planning stages, on how to share data ethically.



Case Study - Victoria University: A journey to an ethics-friendly data-sharing policy

Victoria University (VU), with more than 50,000 students and over 4,500 staff, 'is one of the largest and most culturally diverse education institutions in Australia.' In conjunction with an ANDS-funded project in 2011, the Office of Research began discussions with the VU Manager for Research Ethics and Biosafety, and VU Human Research Ethics Committee and the Animal Experimentation Ethics Committee about data-sharing and ethics. The Manager for Research Ethics sits within the Office of Research making such discussions easier. Previous informal discussions with researchers had shown various misconceptions:

- Ethics committees forbid the sharing of data
- Data must be destroyed after a certain time
- Ethics clearance once granted cannot be amended.

Some ethics committee members suggested in discussion that, contrary to the researchers' views, it is unethical NOT to share data, especially publicly-funded data. Given this feedback, the Office of Research proposed that two new questions be added to their ethics application forms. The questions were:

- Is there any ethical reason NOT to share the data from this project?
- Can the information collected on the ethics form be re-used in other places, such as collecting project descriptions for Office of Research reporting purposes? (eg. Research Data Australia)

The Human Research Ethics committee agreed that:

- There is no *prima facie* ethics requirement to keep data private provided that consent is obtained to share and reuse it, appropriate access mechanisms are in place and approval has been given by the relevant ethics committee.
- There is no requirement to destroy data without good reason. Some Indigenous Australian data is required by statute to be destroyed after a certain time. Current VU ethics documentation is ambiguous about data destruction. Future policy, now under discussion, will require a re-assessment of data after five years to decide if it should be kept or destroyed.
- Ethics clearance, once granted, may be amended by re-application to the relevant ethics committee.
- In some situations, the Human Research Ethics Committee may approve the sharing of data without collecting further consent and under conditions which agree with national guidelines.

Victoria University is paving the way for an ethics-friendly data-sharing policy.



5. Anonymising data

Before data obtained from research with people can be published or shared with other researchers, it may need to be anonymised so that individuals, organisations and businesses cannot be identified from the data.

Anonymisation may be needed for ethical reasons to protect people's identities, for legal reasons to not disclose personal data, or for commercial reasons. Personal data should not be disclosed from research information, unless a respondent has given specific consent to do so.

Anonymisation may not be required, for example, in oral histories where it is customary to publish and share the names of people interviewed and for which they have given their consent.

It can be time-consuming, and therefore costly, to anonymise research data, in particular qualitative textual data. This is especially the case if not planned early in the research or left until the end of a project.

Removing spatial references prevents disclosure, but it means that all geographical information is lost. A better option may be to keep spatial references intact and to impose access regulations on the data instead. As an alternative, point co-ordinates may be replaced by larger, non-disclosing geographical areas or by meaningful alternative variables that typify the geographical position.

Consideration should be given to the level of anonymity required to meet the needs agreed during the informed consent process. Researchers should not presume the only way to maintain confidentiality is by keeping data hidden. Obtaining informed consent for data sharing or regulating access to data should also be considered at the same time as any anonymisation as part of the research planning process.

Data may be anonymised by:

- removing direct identifiers, e.g. name or address
- aggregating or reducing the precision of information or a variable, e.g. replacing date of birth by age groups
- generalising the meaning of detailed text, e.g. replacing a doctor's detailed area of medical expertise with an area of medical speciality
- using pseudonyms
- restricting the upper or lower ranges of a variable to hide outliers, e.g. top-coding salaries

A person's identity can be disclosed from:

- direct identifiers, e.g. name, address, postcode information or telephone number
- indirect identifiers that, when linked with other publicly available information sources, could identify someone, e.g. information on workplace, occupation or exceptional values of characteristics like salary or age

Special attention may be needed for:

- relational data, where relations between variables in related datasets can disclose identities
- geo-referenced data, where identifying spatial references such as point co-ordinates also have data related to those co-ordinates



Managing anonymisation:

- plan anonymisation early in the research as part of your data management planning
- retain original unedited versions of data for use within the research team and for preservation
- create an anonymisation log of all replacements, aggregations or removals made
- store the log separately from the anonymised data files
- identify replacements in text in a meaningful way, e.g. in transcribed interviews indicate replaced text with [brackets] or use XML markup tags <anon>.....</anon>

Audio-visual data

Digital manipulation of audio and image files can be used to remove personal identifiers. However, techniques such as voice alteration and image blurring are labour-intensive and expensive and are likely to damage the research potential of the data. If confidentiality of audio-visual data is an issue, it is better to obtain the participant's consent to use and share the data unaltered, with additional access controls if necessary.

6. Access control and licensing

Under certain circumstances, sensitive and confidential data can be safeguarded by regulating use of or restricting access to such data, while at the same time enabling data sharing for research purposes. It is, therefore, important to consider where and how the data will be managed for the longer term as there need to be systems in place to protect confidentiality and manage access.

Data held at data centres and archives, such as the Australian Data Archive (ADA), is not generally for public use. Its use is restricted to specific purposes after user registration. Users sign an End User Licence in which they agree to certain conditions, e.g. not to use data for commercial purposes or identify any potentially identifiable individuals through data mining or other techniques.

Data centres may impose additional access regulations for confidential data such as:

- needing specific authorisation from the data owner to access data
- placing confidential data under embargo for a given period of time until confidentiality is no longer pertinent
- providing access to approved researchers only
- providing secure access to data through enabling remote analysis of confidential data but excluding the ability to download data

Mixed levels of access regulations may be put in place for some data collections, combining regulated access to confidential data with user access to non-confidential data. Data centres typically liaise with the researchers who own the data in selecting the most suitable type of access for data. Access regulations should always be proportionate to the kind of data and confidentiality involved. Access conditions which require that the data centre contact the researcher directly about each particular request may result in extended delays before access is granted, so it is preferable (but not always possible) for the data centre to be able to deal with the access request.

Research institutions offering facilities for the storage and access of sensitive and confidential data will need to have similar facilities in place to ensure properly regulated access.



Case Study - Aboriginal and Torres Strait Islander Data Archive (ATSIDA)

The Aboriginal and Torres Strait Islander Data Archive (ATSIDA) is a specialist archive for the management of Australian Indigenous Research Data. Based at the University of Technology, Sydney (UTS), ATSIDA is a thematic archive of the Australian Data Archive (ADA), with its datasets stored securely at the Australian National University Supercomputer Facility (ANUSF).

ATSIDA is guided by a set of Protocols that aim to assist with best practice management of the data archive. Formulated with the expert advice of the ATSIDA Reference Group, the Protocols address issues of preservation, access, reuse and return of research data relating to Aboriginal and Torres Strait Islander communities. It is essential that the data archive manages research data that is informed by an appropriate ethics and consent process. ATSIDA works closely with researchers and Aboriginal and Torres Strait Islander communities to manage the complex intersection of rights and interests that affect all of its constituents – researchers, holding institutions, and, most importantly, Aboriginal and Torres Strait Islander communities.

ATSIDA encourages researchers to develop meaningful dialogue with Aboriginal and Torres Strait Islander research participants on setting access and use conditions to data deposited to the archive. Access conditions should be set with the input of participants, and should consider any requirements for managing cultural or confidential information. The moral rights of Aboriginal and Torres Strait Islander contributors should be acknowledged and observed in any reuse of data held in the archive. Conditions of use set by depositors may include attribution of Indigenous knowledge captured in a research dataset.

It is of vital importance that ATSIDA encourages the management of research data in ways that respect local cultural protocols for access to, reproduction and circulation of cultural knowledge. ATSIDA is building a data archive service that aims to strengthen cultures while at the same time contributing to capacity building for Aboriginal and Torres Strait Islander communities. The development of strong and reciprocal relationships between researchers, the Aboriginal and Torres Strait Islander communities and the data archive are a primary focus for ATSIDA in building trust and respect for managing Indigenous research data.

All staff working with the data managed by ATSIDA, or who may come into contact with it through their roles (for example, IT staff), are required to sign and adhere to a confidentiality agreement. Data and associated documentation relating to ATSIDA datasets is managed in accordance with the conditions of access and use set by researchers at the point of deposit. ATSIDA staff are required to manage sensitive and confidential data in a secure and culturally responsive environment.

Additional processes and safeguards are built into ATSIDA processes to manage cultural materials in response to community requests. ATSIDA staff aim to establish effective communication with researchers and Aboriginal and Torres Strait Islander communities about the management of research data to develop a data archive that encompasses the desirable goal of creating a trusted Indigenous data archive.

www.atsida.edu.au



Licensing

There are several ways of providing a restricted licence for data should one be required.

- AusGOAL provides a restrictive licence template which been developed specifically for material that may contain personal or other confidential information and which cannot be covered by a Creative Commons licence.⁶
- The Australian Data Archive (ADA) has forms available which set out data access and use conditions. These forms could be readily adapted for use by other bodies.⁷
- Research institutions may choose to develop their own.

7. For Further Information

Australian Data Archive. *Conditions of access to data from ADA*. <http://www.ada.edu.au/ada/access-conditions>

ANDS: *De-identifying your data*. <http://ands.org.au/resource/data-deidentification.html>

David Lawrence. *Analysis of public use microdata files: A researcher's perspective*. National Statistical Service. 2010.

<http://www.nss.gov.au/nss/home.nsf/NSS/58CC18F435E0800CCA2579E20006BA4B?opendocument>

US Office of Management and Budget, Federal Committee on Statistical Methodology. *Statistical Policy Working Paper 22 (Second version, 2005): Report on Statistical Disclosure Limitation Methodology*. 2005
<https://fcsmsites.usa.gov/files/2014/04/spwp22.pdf>

Acknowledgement

This guide has been adapted with permission from the Ethics and Consent section of *Managing and Sharing Data* produced by the UK Data Archive.
<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

⁶ <http://www.ausgoal.gov.au/restrictive-licence-template>

⁷ <http://www.ada.edu.au/ada/ada-undertaking-forms>

