Australian Research Data Commons

Data sharing considerations for Human Research Ethics Committees

A practical guide to support those on Human Research Ethics Committees to confidently assess applications that propose to share data; understand and promote the processes which ensure safe data management of sensitive data throughout the data lifecycle, and provide advice to researchers about designing their research so the data can later be shared ethically and legally.

Who is this for?

This guide is intended for researchers to use as a supplement to institutional policy regarding ethical research and data sharing.

Last updated: Feb 2020

Contents

The role of Human Research Ethics Committees in data sharing

Human Research Ethics Committees (https://hrea.gov.au) will increasingly need to consider applications that address the intent to share data. There is a changing landscape around data sharing, including:

- The Human Research Ethics Application (HREA) asks several questions related to sharing of data, in particular Q3.13-Q3.18.
- Funder guidelines and requirements include data management and sharing (see Section 3).
- Publisher policies requiring links between articles and data (See Section 8), for example:
    - PLOS journals' policy on making all data available (https://journals.plos.org/plosmedicine/s/data-availability) "with rare exception".
    - The BMJ requires data sharing on request for all trials.
      Note: the data required to be shared is usually the data supporting the results from the publication, and in a de-identified form.
- New capability within data management, which now enables sensitive data to be shared safely.

In research with people, there can be a perceived tension between data sharing and data protection where research data contains personal or sensitive 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 by ensuring that researchers can publish data safely and ethically in regards to the following considerations:

- legislation;
- funder guidelines and requirements;
- publishers' data policies;
- informed consent and data sharing;
- de-identification;
- access control;
- licensing.

As of September 2019, 18 Australian universities mandate Data Management Plans as part of all research projects involving data. During the HREC review process, only the parts of the data management plan relevant to research ethics are needed to be reviewed.

Legislation

Under the Privacy Act 1988 (https://www.legislation.gov.au/Details/C2016C00979), sensitive human and personal data cannot generally be shared in their original form. However, once de-identified, these modified data no longer trigger the Act as they are not 'personal information'. In other words, de-identified sensitive data can legally be shared.

It is worth noting that whilst the Privacy Act 1988 does not apply to de-identified data, it does apply to the activity of de-identifying the data (i.e., removing identifying information from the original, sensitive dataset), and it might also apply in the context of seeking to re-identify data. This activity is, however, explicitly condoned in the Australian Privacy Principles (https://www.oaic.gov.au/individuals/privacy-fact-sheets/general/privacy-fact-sheet-17-australian-privacy-principles) of the Privacy Act 1988 as one of the few exceptions to sensitive data use. This is because de-identification is considered a 'normal… practice' that 'an individual may reasonably expect their personal information to be used or disclosed for' without requiring specific consent.

Australian legislation that may impact on the sharing of sensitive data

- Privacy Act 1988 (https://www.legislation.gov.au/Details/C2016C00979) (Commonwealth and state equivalents): provides definitions of personal information, sensitive information, identification information and de-identified information in Part II, Division I, Section 6.
- Australian Human Rights Commission Act 1986 (http://www.austlii.edu.au/au/legis/cth/consol_act/ahrca1986373) (Commonwealth) and state equivalents.
- Freedom of Information Act 1982 (http://www.austlii.edu.au/au/legis/cth/consol_act/foia1982222) (Commonwealth) and state FOI and Right to Information (RTI) equivalents.

Funder guidelines and requirements

Funder guidelines are increasingly supportive of data sharing:

1. National Health and Medical Research Council (NHMRC) Open Access Policy (https://www.nhmrc.gov.au/about-us/resources/open-access-policy) acknowledges the importance of making data publicly accessible and strongly encourages sharing of data and other research outputs from NHMRC supported research.
2. National Statement on Ethical Conduct in Human Research (https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018), which covers research on human subjects, recognises the value of making data available for future research.
3. Australian Code for the Responsible Conduct of Research has as one of its principles of responsible research: "Transparency in declaring interests and reporting research methodology, data and findings - Share and communicate research methodology, data and findings openly, responsibly and accurately" (P3).
   It also requires Institutions to: "Provide access to facilities for the safe and secure storage and management of research data, records and primary materials and, where possible and appropriate, allow access and reference", (R8) and researchers to "Retain clear, accurate, secure and complete records of all research including research data and primary materials. Where possible and appropriate, allow access and reference to these by interested parties." (R22).
4. The Australian Research Council's (ARC) Research Data Management Strategy states: "The ARC's position reflects an increased focus in Australian and international research policy and practice on open access to data generated through publicly funded research."

International funders

Australian researchers regularly participate in international collaborations, and many of the major international funders of research require sharing of data, e.g.:

- Horizon 2020 statement on FAIR data sharing (https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf) from the EU
- Wellcome Trust (https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing) in the UK
- Medical Research Council (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/data-sharing) in the UK
- National Institutes of Health (NIH) (https://grants.nih.gov/policy/sharing.htm) in the USA
- Bill and Melinda Gates Foundation (http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy) in the USA
- National Science Foundation (https://www.nsf.gov/bfa/dias/policy/dmp.jsp) in the USA

Publishers' data policies

The BioMed Central blog (http://blogs.biomedcentral.com/bmcblog/2016/07/05/promoting-research-data-sharing-springer-nature) explains why publishers are increasingly wishing to strengthen the links between articles and their supporting data.

Whilst there is a wide range of journal data policies, increasingly researchers submitting an article for publication need to indicate the availability of the supporting data. Examples of journal data policies:

- Requiring all data underlying a journal article to be made available with no or minimal restrictions, e.g.
  - PLOS Medicine policy on making all data available (http://journals.plos.org/plosmedicine/s/data-availability) “with rare exception”
  - Nature (http://www.nature.com/authors/policies/availability.html)
  - PNAS (Proceedings of the National Academy of Sciences USA) (http://www.pnas.org/site/authors/journal.xhtml)
- Requiring a statement on the authors' willingness to share the data, e.g.
  - Annals of Internal Medicine (http://annals.org/public/authorsinfo.aspx#data-sharing-and-reproducible-research)
  - The BMJ requires data sharing on request for all trials (http://www.bmj.com/content/350/bmj.h2373)
  - The American Political Science Association has been fostering a Data Access & Research Transparency (DA-RT) (http://www.dartstatement.org) program for political science, including integrating the DA-RT principles into their Ethics Guide (http://www.dartstatement.org/2012-apsa-ethics-guide-changes).

Hear it from the experts: Clinical data disclosure in 90 seconds (YouTube, 90 sec - https://www.youtube.com/watch?v=DI08xpfOVf0&list=PLG25fMbdLRa5pvodHMYDi3c0LTu8N3Ks-&index=6) with Iain Hrynasckiewicz, Head of Data and HSS Publishing, Open Research Nature Publishing Group & Palgrave Macmillan.

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. Personal data should never be disclosed, unless consent has been given for disclosure. Personal and/or sensitive data can be shared if consent has been obtained and if suitable procedures, precautions and safeguards are followed. Even if consent is provided for sharing personal data, researchers still need to exercise judgement to preserve the interests of participants and, in the event that a particular risk is identified that the participant may not have considered, that judgement should err on the side of NOT sharing the data if that risk is material and would, reasonably, have caused a participant to decline consent to disclose personal information (e.g. disclosure might trigger legal action against the participant).

Consent documentation should:

- avoid precluding de-identification, publication and sharing of data
- inform participants how research data will be stored, preserved and used in the long-term
- inform participants how privacy will be maintained, e.g. by de-identifying data and/or restricting access for secondary use to legitimate researchers
- state the conditions under which access to the data may be granted to others
- obtain explicit informed consent for data sharing
- refer to information that describe any risks related to how the data might be used.

Consent documentation should contain:

- the level of consent. The National Statement on Ethical Conduct in Human Research (https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018) gives three levels of consent for the future use of data; specific, extended or unspecified (Section 2.2.14). Whichever one is chosen by the researchers must be made clear to the research participants.
- explicit information on whether the data is to be held in a form which is identifiable, non-identifiable or re-identifiable. For more information see the ARDC guide on Data de-Identification (http://www.ands.org.au/working-with-data/sensitive-data/de-identifying-data).

Wherever possible, the value of the data to the wider research community should be taken into account during the research planning process, as should data preservation and longer-term use. Patient information sheets and consent forms should be designed accordingly. At a minimum, they should not preclude data sharing, such as by promising to destroy data unnecessarily. The practice of destroying data when there is not a good reason to do so does not recognise that data collection imposes a significant burden on participants, and destruction may trigger requests for new or further data collection that imposes a potential 'I agree that research data gathered for the study may be published provided my name or other identifying information is not used.'

Example sentences for consent forms to request data publication and sharing

Where data is intended to be public or accessed with little restriction

'The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your identity as a participant in this study is confidential.'

'Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.'

'I agree that research data gathered for the study may be published provided my name or other identifying information is not used.'

For data that will have conditional access

'Other genuine researchers [may] have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.'

The above example may be adapted to include specific access conditions that you intend to apply to its reuse:

'Other genuine researchers may request access to de-identified data in the future. Access will only be granted if they agree to preserve the confidentiality of the information as requested in this form. Their access will also require approval from the original research team as well as approval from a Human Research Ethics Committee at their home institution.'

Read Recommended informed consent language for data sharing from the ICPSR (https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/conf-language.html), which has been managing large quantities of sensitive Social Science data in the USA for over 50 years.

Researchers may also consider giving participants the opportunity to select whom they agree to share their data with (and whom they don't) such as from a list of likely data re-users. As an example, in the case of data in Indigenous studies being archived e.g. by AIATSIS (http://aiatsis.gov.au), participants can elect to allow access to specified individuals (e.g. family members).

Sharing existing data, when re-contact with research participants is not possible

Data, including sensitive data can be shared can be shared without explicit consent from research participants if:

1. The information given to participants prior to their consent for data collection indicated future use of the data*,

   OR

2. The opportunity to gain consent no longer exists or is not practical, and
   - The data can and have been de-identified, and
   - The process of de-identification matches the definition provided in the Privacy Act 1988 (https://www.legislation.gov.au/Details/C2016C00979), and
   - There is no risk that publishing or sharing the data will cause harm or contribute to discrimination towards the research participants or subjects, and
   - Information Sheets and Consent forms from the original data collection did not preclude sharing.

* In cases where participant Consent Forms did not refer specifically to data publication or sharing (though not precluded it either) and Information Sheets did, consent to participate in the project itself allows sharing. This is because informed consent implies an understanding and agreement to the Information Sheet.

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 de-identification as part of the research planning process.

De-identification

Before data obtained from research with people can be published or shared with other researchers, it may need to be de-identified (https://ardc.edu.au/resources/working-with-data/de-identification) so that individuals, organisations and businesses cannot be identified from the data.

There is also the need to assess the risk of re-identification that may occur when other parties hold data assets that enable individuals to be re-identified when two or more data sets are linked. The Five Safes Framework (https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017) provides a structure for assessing and managing disclosure risk that is appropriate to the intended data use.

The Privacy Act 1988 (https://www.legislation.gov.au/Details/C2016C00979) defines de-identified as "personal information is de-identified if the information is no longer about an identifiable individual or an individual who is reasonably identifiable" (Part II, Division I, Section 6).

Personal data should not be disclosed, unless a respondent has given specific consent to do so, this may contravene the Privacy Act 1988 (https://www.legislation.gov.au/Details/C2016C00979). For example, de-identification may not be required in oral histories where it is customary to publish and share the names of people interviewed and when they have given their consent to publish.

For further information:

- see the ARDC guide on De-identification (https://ardc.edu.au/resources/working-with-data/de-identification)
- Australian Bureau of Statistics Managing the risk of disclosure: The Five Safes Framework (https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017)
- Australian Office of the Information Commissioner: De-identification decision making framework (https://www.oaic.gov.au/privacy/guidance-and-advice/de-identification-decision-making-framework)

Access control

	Open	Mediated / controlled access	Closed
Discoverability	Metadata fully discoverable	Metadata fully discoverable	Metadata not publicly available
Accessibility	Data accessible and immediately downloadable	Mediated access to data via data custodian. • May be de-identified • Conditions around who can access data for what purposes	Data not discoverable or available
Sensitivity	Non-sensitive data from completed projects	Sensitive data from completed projects	Highly sensitive data (e.g. commercial in confidence or national security OR data form projects not yet completed

There is capability within data management that enables sensitive data to be shared safely. Sensitive data can be safeguarded by regulating the use of, or restricting access to, such data (known as mediated or controlled access), while at the same time enabling data sharing for further research purposes and/or for replication when the research findings are published in journals.

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 manage access. The Five Safes System (https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017) offers data custodians a framework to place appropriate controls, not just on the data itself, but on the manner in which data are accessed. The framework is designed to facilitate safe data release and prevent over-regulation.

Access control in data archives and data centres

There are data archives available that can provide appropriate access controls and secure data storage. Data held at data archives, such as the Australian Data Archive - (ADA) - (http://www.ada.edu.au) and the UK Data Archive (UKDA) - (http://www.data-archive.ac.uk), is not generally for public use. Metadata (https://ardc.edu.au/resources/working-with-data/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 and application.

- Data centres may impose additional access regulations for sensitive data such as:
- needing specific authorisation from the data owner to access data
- placing sensitive data under embargo for a given period of time until sensitivity is no longer pertinent
- providing access to approved researchers only
- providing secure access to data through enabling remote analysis of sensitive data in a secure environment, but excluding the ability to download data
- requiring data to be de-identified unless consent has been given for personal information to be shared.

Access control in research institutions

Research institutions offering facilities for the storage and access of sensitive data should consider having similar facilities in place to ensure properly regulated access. Access to data can be managed by a data custodian within a research institution. The data custodian could be the original researcher, a staff member or researcher from the Faculty or School, or from the repository where the data is deposited. Ideally there would be a system in place to assist with decision making around applications for access to the data.

Many information and consent sheets indicate that only members of the research team will have access to data, and such statements potentially preclude access by a third-party data custodian. It may be useful to include relevant IT staff or data custodians among the list of those who might have access to sensitive data; such personnel are typically bound by institutional policies around privacy and security.

Licensing

All Australian data intended for reuse should have a license. A license is a document that clearly sets out how the data can be used and attributed to the original data owner. The ARDC's Research Data Rights Management Guide (https://ardc.edu.au/resource/research-data-rights-management-guide) can help researchers and data owners choose the appropriate licence to be added to data.

It is important to note that applying a license to data does not allow a researcher to publish sensitive data, or act as a substitution for data de-identification. Sensitive data remains sensitive even with a license, and thus cannot be published without participant consent, and de-identification if appropriate.

Licensing within data repositories

Researchers using data from, or depositing data in, a data repository usually sign an End User License (e.g. ADA (http://www.ada.edu.au/ada/access-conditions), UKDA (https://www.ukdataservice.ac.uk/get-data/how-to-access/conditions#/tab-end-user-licence) in which they agree to certain conditions, for example not to use data for commercial purposes or identify any individuals through data mining or other techniques.

For more detailed information on Licensing see:

- Licensing for data reuse
- Research data licensing data and copyright FAQ

FAIR

The FAIR (https://ardc.edu.au/resources/working-with-data/fair-data) principles for data management and sharing are gaining momentum internationally. For instance, the European Code of Conduct for Research Integrity Revised Edition 2017 (https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf) has explicitly endorsed the FAIR principles:

> "Researchers, research institutions and organisations ensure access to data is as open as possible, as closed as necessary, and where appropriate in line with the FAIR Principles (Findable, Accessible, Interoperable and Re-usable) for data management." (p.6)

In the Australian context, as stated in the NHMRC Guide to the Management of Data and Information in Research (https://www.nhmrc.gov.au/sites/default/files/documents/attachments/Management-of-Data-and-Information-in-Research.pdf) - a guide supporting the Australian Code for the Responsible Conduct of Research):

> "Published research data generally... should be findable, accessible, interoperable, and re-usable, both manually and with automated tools."

Conclusion

Sensitive data can be shared ethically and legally if researchers consider four important aspects:

1. including provision for data sharing when gaining informed consent
2. protecting people's privacy by de-identifying data, and assessing the risk of re-identification, where needed
3. considering controlling access to data
4. applying an appropriate license.

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. This is particularly important in the context of a changing landscape amongst institutions, funding agencies and publishers, a great number of whom now explicitly encourage and even in some cases require data to be shared. Such requirements may not have been envisaged at the outset of a research project, and inadequately preparing for the prospect of data sharing may seriously limit researchers' access to funding and publishing opportunities in the future.

About the Australian Research Data Commons

The Australian Research Data Commons (ARDC) is a transformational, sector-wide initiative, working with sector, government, and industry partners to build a coherent national and collaborative research data commons. This will deliver a world-leading data advantage, facilitate innovation, foster collaboration and enhance research translation.

Visit ardc.edu.au for more information.

Contact us

Call us +61 3 9902 0585
Head to our site ardc.edu.au
Tweet us @ARDC_AU
Guide ardc.edu.au/guides/data_sharing_considerations_for_hrecs

The Australian Research
Data Commons (ARDC)
is enabled by NCRIS.

NCRIS
National Research
Infrastructure for Australia
An Australian Government Initiative