Publishing Sensitive Data

Updated 9/09/2022

Key messages from the ARDC Guide to Publishing Sensitive Data

- The benefits of publishing your sensitive data and metadata can be significant as long as appropriate steps are taken to meet your ethical and legal obligations.
- Publishing your data and metadata means that others can discover it and cite it. Increasing the visibility of your data this way can lead to new collaborations, improve the impact of your research, and create opportunities for authorship.
- You can publish a description of your data (i.e. metadata) without making the data itself openly accessible.
- You can place conditions around access to published data.

CONTENTS

INTRODUCTION

About this guide

This Guide outlines best practice for the publication and sharing of sensitive research data in the Australian context. The Guide follows the sequence of steps that are necessary for publishing and sharing sensitive data, as outlined in the Publishing and Sharing Sensitive Data Flowchart. It provides the detail and context to the steps in this poster. References for further reading are provided for those that are interested.

By following the sections below, and steps within, you will be able to make clear, lawful, and ethical decisions about sharing your data safely.

How the Guide interacts with your institutional policies

This Guide is intended to support institutional policies on data management and publication. Most researchers operate within the policies of their institution and/or funding arrangements and must, therefore, ensure their decisions about data publication align with these policies. This is particularly relevant for Intellectual Property, and sometimes, your classification of sensitive data (e.g. NSW Government Department of Environment & Heritage, Sensitive Data Species Policy) or selection of a data repository.

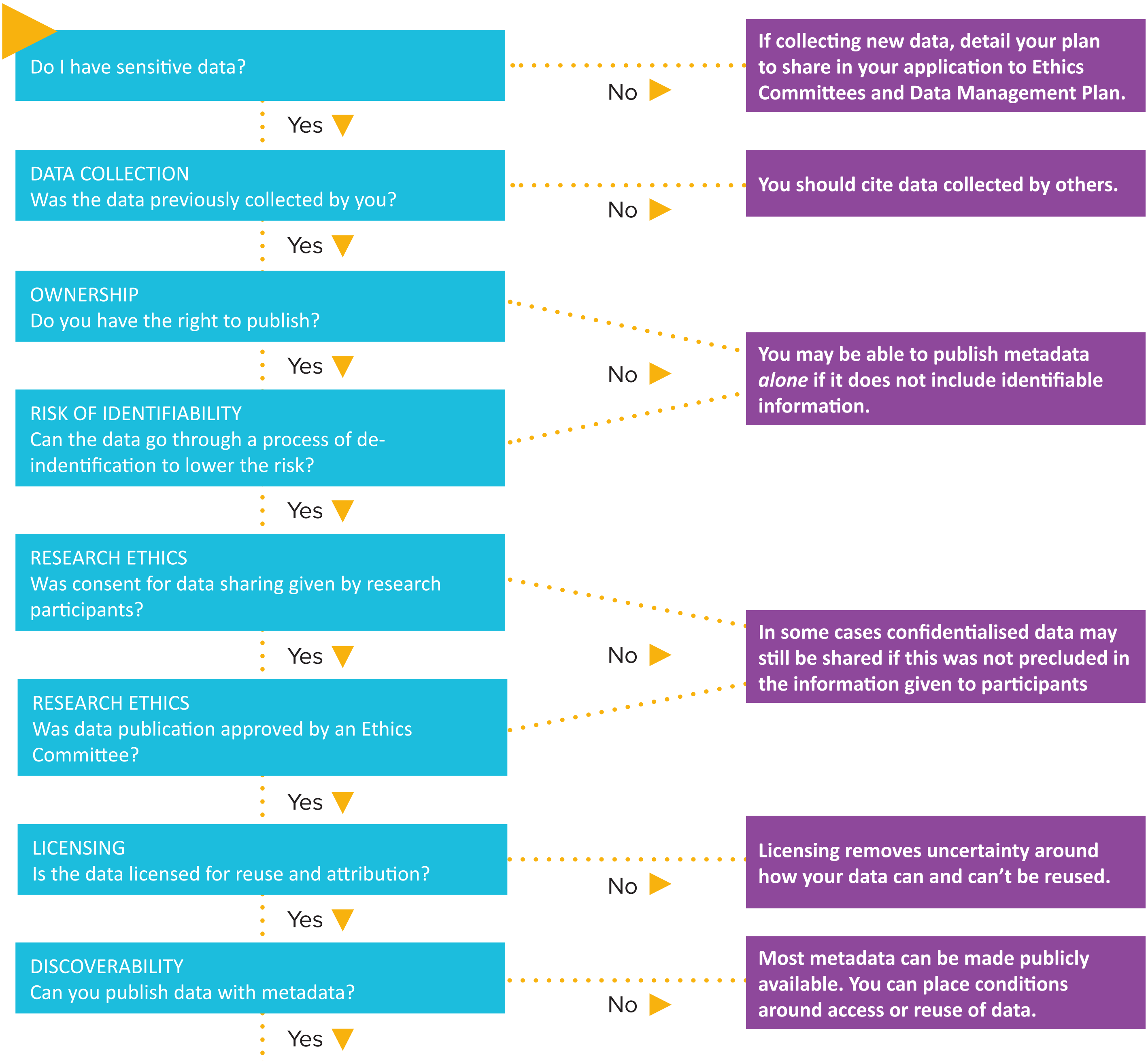
Publishing and sharing sensitive data

Australian Research Data Commons

When and how to publish sensitive data as openly and ethically as possible.

For more information see: ardc.edu.au/resources/working-with-data

Sensitive data identifies individuals, species, objects or locations, and carries a risk of causing discrimination, harm or unwanted attention.



Do I have sensitive data?

No ▶ If collecting new data, detail your plan to share in your application to Ethics Committees and Data Management Plan.

Yes ▼

DATA COLLECTION
Was the data previously collected by you?

No ▶ You should cite data collected by others.

Yes ▼

OWNERSHIP
Do you have the right to publish?

Yes ▼

RISK OF IDENTIFIABILITY
Can the data go through a process of de-indentification to lower the risk?

No ▶ You may be able to publish metadata *alone* if it does not include identifiable information.

Yes ▼

RESEARCH ETHICS
Was consent for data sharing given by research participants?

Yes ▼

RESEARCH ETHICS
Was data publication approved by an Ethics Committee?

No ▶ In some cases confidentialised data may still be shared if this was not precluded in the information given to participants

Yes ▼

LICENSING
Is the data licensed for reuse and attribution?

No ▶ Licensing removes uncertainty around how your data can and can't be reused.

Yes ▼

DISCOVERABILITY
Can you publish data with metadata?

No ▶ Most metadata can be made publicly available. You can place conditions around access or reuse of data.

Yes ▼

PUBLISHING AND SHARING
Keep in mind

- Make your data as FAIR as possible (Findable, Accessible, Interoperable and Reusable).
- Publish your data and metadata according to participant consent, ethics approval and licensing.
- For sensitive data, it is often appropriate to have public metadata and conditional access to the data whose risk of identifiability has been managed.
- Cite your data along with your other scholarly outputs.

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

The Australian Research Data Commons (ARDC) is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy.

Why publish and share sensitive data?

Box 1. Quick Definitions

Data publication versus sharing: Publication occurs when data is made public. This includes having a publically-available description of the data and access, or information about conditional access, to the data itself. Data sharing occurs when data is made available to others, but does not always accompany publication (e.g. when data is shared among colleagues but not publically discoverable or available).

Metadata: data about your data; i.e. a description of your data. A common example of metadata is a library catalogue record.

Repository: a place where metadata records and — often but not always — data are stored. Data repositories usually have an online portal where members of the public can search for and discover data. There are discipline-specific, institutional-specific, and general repositories.

De-identified data: data that has been modified to reduce the risk that people or subjects of the data can be identified.

Australia and other nations have significant and high-quality datasets that contain potentially sensitive information. This includes epidemiological surveys of health, medical trial data, and ecological studies of endangered species. For reasons of finance, efficiency, conservation, and participant fatigue and disturbance, these should be reused rather than repeated where possible.[1, 2] Sharing this valuable data not only improves the transparency of research, but can increase the impact of research and the benefits to the wider community, thereby improving the return on investment into research. In recognition of this, there have been increasing requirements from publishers and funding bodies for researchers to publish and share their data.[3, 4, 5, 6]

The advantages to publishing and sharing data go beyond meeting publisher or funder requirements. The benefits to the researcher, institution and research participants are clear.

- If data or a description of a dataset are published they are discoverable by others, and can thus be cited and attributed to the original data collector or owner.
- There is evidence that scholarly papers that are accompanied by published data are cited more often than those without.[7]
- New collaborations and publications may eventuate if you choose to share your data with others for reuse.

- Making data available allows re-analysis of expensive, rare or unrepeatable investigations.
- Sharing data from medical studies reduces unnecessary duplication and maximises use of data from these studies, thus reducing the burden upon participants.[8]
- Storing your data in a public repository affords secure and ongoing storage that may not be available in your current or changing research environment.

WHAT IS SENSITIVE DATA?

Defining sensitivity

Sensitive data are data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions.

Data about people

Human data is a common form of sensitive data. Sensitive human data most commonly refers to sensitive personal information, although other types of data to do with people may also be sensitive (for example, sensitive cultural data)

Personal information is information that can be used to identify a person or group of people.[9, 10, 11] You should be careful when considering publishing any form of personal information, as the disclosure of personal information is typically controlled by Australian privacy legislation. Different legislation applies in different jurisdictions.[12] When in doubt, consult your institution's privacy officer.

Some personal information carries a particular risk of discrimination, harm or unwanted attention for the individuals it identifies. For example, the Privacy Act 1988 states that personal information is sensitive when it directly identifies a person (e.g., name, date of birth, address) and accompanies one or more pieces of information from Table 1. 'Sensitive' personal information is subject to tighter controls under this legislation.

Table 1. Types of sensitive information, as defined by the Privacy Act 1988 *(Part II, Division I, Section 6)*

- racial or ethnic origin
- political opinions
- membership of a political association
- religious beliefs or affiliations
- philosophical beliefs
- membership of a professional or trade association
- membership of a trade union
- sexual orientation or practices
- criminal record
- health information (see section 6FA for the definition of health information)
- genetic information
- biometric information

It is important to note that removing obvious identifiers like names and addresses does not necessarily mean that the possibility of identifying an individual has been removed. Information about when, how, and where a dataset was collected may increase the identifiability of the research participants. Combining different kinds of information and data to identify individuals or groups is referred to as 'triangulation', and the current information age means that there is more data in the world than ever before.

There are particular considerations that must be made around sharing data from or about Aboriginal and Torres Strait peoples and practices. Specialised information about sharing these kinds of data can be found in:

- AIATSIS Collection Access and Use Policy
- AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research
- Guide to applying the AIATSIS Code

Non-human sensitive data

Human data is not the only kind of data that can be sensitive. Under the definition given above, sensitive data includes any data that could introduce a risk of harm. Non-human sensitive data includes ecological data, commercial data, and data that poses a threat to national security. Because data can be sensitive in many different ways, it is not possible to give exact instructions for how to handle all types of sensitive data. However, considering the kind of harm that could be posed by your data, the severity of that harm, and its likelihood will help you to work out if and how you can share your data. Your institution may have a tool or guide for classifying the sensitivity of your data. There may also be regulations and discipline standards that can guide you. Here we will consider the example of sensitive ecological data.

Check: Are your data sensitive?

If your data contains information from Table 1 or information about secret or sacred practices, or information that would result in an adverse effect on a species if made public, it is likely to be sensitive.

In addition to looking at the bare content of the data, ask whether, if shared, the data could potentially cause harm or contribute to discrimination to determine if the data are sensitive.

An illustration of sensitive ecological data

Sensitive ecological data includes data that reveals the location of rare, endangered or commercially-valuable species, or other conservation efforts.[13] The Australian Government Department of the Environment[14] notes that the degree of sensitivity needs to be considered in the light of the type of threat, level of threat, vulnerability to the threat, and the type of information. Definitions of ecological sensitivity are not clear-cut, in part due to the sometimes-transitional nature of risk of harm in this field. For example, a population of frogs may be small and declining in one location but not another, and thus, considered sensitive in the former but not latter.

Chapman and Grafton (2008)[15] at the Global Biodiversity Information Facility (GBIF) define sensitive data as information that 'if released to the public, would result in an 'adverse effect'' to the species or conservation activity. This GBIF report recommends that data owners clarify which elements of the dataset trigger this sensitivity, and clearly document that decision in the metadata for the dataset.

A survey was conducted in an area of New South Wales over a six-month period to estimate the population and breeding habits of several bird species. The data resulting from this survey included georeferencing data and breeding status of each species. One of the species included in this survey was the Glossy Black Cockatoo, which is listed as a 'vulnerable' species by the NSW Department of Environment & Heritage in their Sensitive Species List. This list states that the reason for the cockatoo's vulnerable status is 'risk of egg collection and nest disturbance'.

Identifying the causes of a species' 'sensitivity' also enables the data owner to isolate which elements of the dataset may be removed or modified to de-identify the dataset (see Ethics and Consent below).

Sensitive by context

What is considered to be sensitive may differ with time and across groups of people or subjects. Non-sensitive data may become sensitive by its context, or when more information is added. The names of an organisation's or publication's subscriber list are not usually sensitive. But this list may be sensitive if the special interest of that group could expose subscribers to discrimination[11]: for example, an extreme or unpopular political affiliation. For ecological data, a species may be at risk of harm from human activity in one geographic area but not another. In this latter case, sensitivity might be determined by the location that the data refers to rather than the species itself.

There are instances where data that is not obviously sensitive (i.e., does not include direct identifiers such as names or dates of birth), or has been modified to reduce their sensitivity, may become sensitive again when context changes. Two common examples are:

1. Triangulation: When the identity or sensitivity of a participant or subject can be determined by combining several pieces of non-sensitive information. For example, with human data, if you have information about a person's age, occupation and family composition, it may not be difficult to identify them in a relatively small sample.
2. Data Linkage: Bringing together two or more datasets that include the same person or subject of research. Alone, a dataset may not contain enough information to identify individuals or place subjects at risk, but when combined with two or more datasets, this may now be achievable.

For example, Dataset A describes the medical history of a group of non-identifiable patients with a cognitive disorder. Dataset B includes employment information and public transport usage in the same, moderately small population. When these datasets are linked, there could be sufficient information about where patients work and live such that they may be identified.

An example from ecological research: Dataset A describes a species of falcon that is vulnerable to egg collection, and its population and location over time. Dataset B includes breeding patterns of the same falcon species over a similar period of time. When linked, there may be enough information to determine the location of falcons at nesting time.

Data owners and managers should always consider the possibility for triangulation in their dataset and check for this. It is good practice to reconsider triangulation with the introduction of new data; that is, following data linkage.

I HAVE SENSITIVE DATA - NOW WHAT?

How to publish and share sensitive data

The previous section describes sensitive data in its original form. In some cases, it may be possible to modify the data to reduce the risk that it poses. For human data, this might mean reducing the identifiable information present in the dataset so that it is much less likely that individual participants can be identified. There are many methods for achieving this - see the ARDC's Identifiable Data resources. This kind of modified data is still highly valuable and re-usable to some other researchers, and the data can in some cases be published and shared with fewer legal and ethical constraints.

It is important to remember that while the risk posed by sensitive data can often be substantially reduced, the data cannot always be assumed to be completely "safe", and there is a possibility that the risk may increase again if circumstances change (see Sensitive by Context, above). For this reason, even if it has been modified as needed to allow sharing, most sensitive data is published with conditional access. This means that a description (i.e., 'metadata') is published in a data repository, and the data itself is accessible once some conditions (set by the researcher and/or the repository) are met. The discoverability of the data is not restricted - i.e., anyone can find and read the metadata. This public description of the data includes information about access and a link to how/where to apply for access.

Legal restrictions on sharing sensitive data

The sharing of certain kinds of sensitive data is controlled by legislation. For human data, one of the most important types of legislation to consider is privacy legislation. Different privacy legislation applies in different Australian jurisdictions[12] - here we will consider the example of the Commonwealth *Privacy Act 1988*[9](Privacy Act).

This Act states that sensitive personal information cannot generally be shared in its original form. However, once de-identified, the modified data no longer triggers the Act. The Privacy Act defines de-identified this way: "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). The Privacy Act also provides a definition of 'identification information', indicating what information must be removed, at a minimum, for individuals to not be identifiable in the data. See "Risk of Identifiability" below for more information about managing the identifiability of data.

It is worth noting that whilst the Privacy Act 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). This activity is, however, explicitly condoned in the Australian Privacy Principles as one of 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[16]. For more information about human consent for data sharing see Ethics and Consent, below.

Remember that the Privacy Act is not the only legislation that may control the sharing of your sensitive data. Consider the privacy legislation that may apply in your state or territory, as well as other relevant legislation. Your institution's privacy officer and legal office can help you with this.

RISK OF IDENTIFIABILITY

When sharing data, it is important to prevent participants or subjects of research (e.g., animal or plant species) from being identified and placed at risk of harm or discrimination. This involves removing or modifying information in the original (sensitive) dataset. The exact information that needs to be removed or modified will vary depending on the contents of the dataset and the reason that the data has been deemed sensitive.

Check: can my data be de-identified?

Some data cannot be modified to sufficiently minimise the risk of harm to participants or subjects without rendering the data largely invaluable. This is often because the context in which the data were collected makes recovery of identity possible. For example, data from in-depth personal interviews, and some kinds of cultural or historical data for which identity is an intrinsic aspect of the data. In the former case, data may also reveal identities of other individuals who did not participate (and thus consent) in the study.
In the majority of cases, data that cannot be de-identified cannot be published without explicit consent from the participants and approval of a Human Research Ethics Committee. This does not necessarily preclude publication of a metadata record if that record cannot be used to identify the persons or subjects involved in the study.

For further information on the process of de-identification, see the ARDC's Identifiable Data resources.

Modifying sensitive data that is not about individuals

The methods by which sensitive data must be modified before it can be shared may be different depending on what it is about the data that makes it sensitive. When sensitivity relates to the identification of individuals, the identifying information must be removed to reduce that risk. Our example of ecological sensitive data requires a different approach. The sensitivity of ecological data typically does not relate to the naming of the species itself, but to accompanying information about its location and/or dates for breeding, fruiting, or migration. For example, the illustrative dataset described above is not sensitive simply because it includes data from the Glossy Black Cockatoo, but because it is accompanied by data about its precise location at breeding time.

Worked examples of ecological data that have been modified to manage sensitivity can be found in Table 1 of the Australian Government Department of the Environment's 'Sensitive Ecological Data – Access and Management Policy V1.0'[14].

Managing the risk of re-identification

Once data has been de-identified, the risk of re-identification must be reviewed if these data are linked with other data. Data linkage is the merging of two or more separate datasets that contain data from the same people or subjects. It is becoming increasingly common in epidemiology, medical, social and ecological sciences because it enables researchers to understand people's or subjects' context in more detail by adding more information without having to collect new data. Data linkage also derives greater value from existing datasets.

Like triangulation, data linkage can mean that de-identified participants or subjects can become re-identified because one or more pieces of potentially-identifiable information have been added (by the newly linked data). This possibility must be assessed when the data are linked by treating the new, linked dataset as an identifiable dataset and assessing the risks.

ETHICS AND CONSENT

In addition to meeting legal and community standards, researchers have ethical obligations towards participants and research subjects. These include preserving privacy and avoiding any possible harm arising from participation in research and its subsequent publication. An important concept within research is consent - that participation in research is the result of a choice made by participants, voluntarily and with sufficient information and understanding. The ethical management of data must be the primary concern of researchers to maintain participants' trust and research integrity.

For new human data, or when contact with research participants is possible

In most situations consent is required from human participants before obtaining and publishing data. The best time to obtain consent is before the data are collected. This is not only best practice, but avoids the expense, delay and loss of use of data that may be occasioned by attempting to obtain consent later in the research process. Concerns that participants will refuse to participate in research if data sharing is requested are likely to be unfounded.[17] [18]

The request for participants' consent to publish and share their de-identified data should involve:

1. Including information about the processes of data de-identification, publication, and sharing in the 'information sheet' that is provided to people before they agree to participate in the research study and before they are asked to consent to data collection. This information must be understandable to the participant and have sufficient detail such that they can make an informed decision about their consent to publish and/or share the data they provide. Participant information sheets must be approved by your Human Research Ethics Committee/s (HREC) before any aspect of the study commences. It should (briefly) cover: procedures for maintaining confidentiality, data publication, and the conditions under which data sharing can occur, including whether there is a prospect of sharing the data with researchers outside those conducting the study.[1]
2. Specifically requesting consent for data sharing and/or publication in the 'consent form', which must also be approved by your HRECs before research begins. Example wording to request data publication and sharing in consent forms are provided below.
3. Familiarise yourself with any requirements for ethics applications, information sheets and consent forms in your jurisdiction with regards to data publication and sharing. For example, the National Health and Medical Research Council’s Human Research Ethics Application (HREA) asks several questions related to sharing of data, in particular Q3.13-Q3.17.

Consent forms requesting data publication and sharing

Statements about data publication and sharing in participant consent forms should:

1. Avoid unnecessarily precluding data de-identification, publication, and sharing[19]
2. State the possibility of future data publication (including storage in a repository) and sharing
3. State the conditions under which access to the data may be granted to others. This may include the process of de-identification as well as other possible conditions, such as approval by the original research team (see Conditional Access, below). In some cases, it may also be appropriate to provide an opportunity for participants to select whom they agree to share their data with (and whom they don't).
4. Be documented with the collected data so subsequent users of the data are aware of the conditions agreed to by participants.[1]

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 participation in this study is confidential.'[20]
>
> 'Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.'[18]
>
> '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.'[21]

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'

You may also consider giving participants the opportunity to select whom they agree to share their data with (and whom they don't). For example, from a list of likely data re-users.

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

In some cases, it may be possible to share sensitive data without the explicit consent of the participants.

It might be the case that you sought consent to collect the data, but failed to ask for consent to future sharing. In this case, consider the *National Statement on Ethical Conduct in Human Research (2007)*, which states that sensitive data can be shared without explicit consent from research participants if:[1]

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, **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.

NB: * 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 consent implies an understanding and agreement to the Information Sheet.

If you did not originally gain consent for future use of the data, can no longer reasonably gain that consent, and believe that the exception applies, you should seek a waiver of the requirement for consent from the appropriate body at your institution - for example, your local Human Research Ethics Committee.

It could instead be the case that you have data for which no consent was obtained in the first place. For instance, your data might consist of public health information. In order for you to use this data in your research without consent, a Human Research Ethics Committee must determine that your use of the data has benefits that outweigh the protection of the privacy of the individuals represented in the data. The same process is required if you want to go on to publish or share this data. You can read more about how these determinations are made in the *Guidelines approved under Section 95A of the Privacy Act 1988.*[22]

Recommendation

If you cannot gain consent to publish the data, you should seek the approval of an appropriate ethics body, such as your local Human Research Ethics Committee.

WHO CAN FIND AND ACCESS THE DATA?

Data can be reused for new and valuable purposes, and potential re-users may exist outside your own professional networks. There could be future applications of the data arising from your research that you would never have thought of.

For data to be reused and cited, other people need to be able to find it and refer to its location. A public metadata record in a repository is the best and easiest way to make data discoverable by others. By doing this, you are still taking the first steps to making your data FAIR. This is a description of the data in a public catalogue. By providing good metadata, you make it easier for others to know if the dataset could be useful for their research.

You can publish a description of data (metadata) without making the data itself freely available

Your chosen repository will provide instructions about what metadata are required before publication. This will almost always be information that you already have in grant applications, project reports, or articles. The metadata record also provides the potential re-user with information about how the data can be accessed - either with a direct link to the data, or instructions about other access arrangements.

Conditional access to data

Sensitive data does not have to be openly available for anyone to access. Conditional access occurs when a metadata record is available to the public (i.e., published in a public repository) but access to the data itself occurs only after predetermined conditions are met. These conditions are set by the researcher or data owner, and/or the data repository. They may include requiring the potential data reuser to:

- Register and/or provide contact details
- Provide information about how they will use, store, or manage the data
- Agree to conditions of data security, privacy
- Agree that they may be contacted by the data owner for purposes of collaboration or otherwise
- Pay an access fee
- Meet other conditions included in Consent Forms and Information Sheets agreed to by original (human) research participants.

Reasons you might want to make your sensitive data discoverable but with conditional access include:

- To control the risk of future reidentification

- To ensure reusers are genuine researchers
- To ensure reusers are aware and agree to maintain confidentiality or secure storage of the data
- You specified conditional access in the participant consent form or Research Ethics Application form
- You would like to maintain some oversight over who uses the data, or for what purpose they are using the data

Other reasons for conditional access include:

- The data (or some of the data) are under embargo
- You would like to be informed about who uses the data so you can collaborate

An example of published metadata with conditional access to data: The Australian Longitudinal Study of Women's Health (ALSWH)

“Data sharing is fundamental for ALSWH as a national research resource to strengthen the evidence base for supporting development of women's health policy and practice. We are fully committed to making our data available and encouraging collaboration between researchers in wide-ranging fields. Data sharing among multidisciplinary groups provides the opportunity for fresh perspectives and for gaining new insights and knowledge on women's health." Professor Gita Mishra, ALSWH Director

Access conditions for the archived Australian Longitudinal Study on Women's Health data in the Australian Data Archive (data repository) are clearly explained in the metadata record.

> “Data may be made available to collaborating researchers where there is a formal request to make use of the material. Permission to use the data must be obtained from the Publications, Analyses and Substudies (PSA) Committee of the ALSWH. Additionally, it is a requirement that a member of the core ALSWH team (a Steering Committee member) be involved as a collaborator on every analysis.“

PUBLISHING THE DATA

Licensing

All Australian data intended for reuse should have a licence. This includes datasets that you have modified for publication. A licence is a document that clearly sets out how the data can be used and attributed. Without a licence, it is unclear how the data can be reused and this may discourage the potential re-user or even lead to accidental misuse.

Licences come in varied forms, ranging from few to many restrictions on reuse. Some data repositories have their own licensing documents (e.g., Australian Data Archive's Access Categories). Others require open access; that is, non-restricted access (e.g. Dryad). If you have very specific requirements about how your data can be reused (frequently the case with sensitive data), you may need to specify your own terms and conditions rather than using a standard licence. Often this is done by entering into a contractual data sharing agreement with the entity which is requesting the data.

It is very important that you consider the implications of applying a particular licence (such standard repository licences) to the data. Do the stated conditions match up with how this particular dataset should be treated? Sensitive data remains sensitive even with a licence. Applying a license does not change your responsibility to make sure that the data is published and shared in a way that is legal and minimises risk. Accordingly, this Guide recommends that you apply a licence, publish, and share the data after the identifiability of the data and other risks have been carefully considered and appropriate measures have been taken. If a particular licence is not appropriate for the data, then you must choose a different licence (even if it means that you have to use a different repository).

For more detail see the ARDC Research Data Rights Management Guide.

How do I know if I own the copyright for the data?

Ownership of copyright and instances of copyright waiver differ across Australian institutions. As copyright is an aspect of intellectual property, we recommend that you look up the Intellectual Property Policy of your institution or employer. If still unclear, seek advice from your Research Office or Research Services Division.

For more detailed information see the ARDC Research Data Rights Management Guide.

Where should I publish the data?

There are many data repositories to choose from. Some repositories provide a catalogue of metadata only and link to, or reference, the storage location of the data, e.g. Research Data Australia. Others both catalogue metadata and store the data themselves e.g. Australian Data Archive, Figshare. Repositories can also be institution-specific, discipline- or content-specific, or general. You can find a list of international repositories at re3data.org.

When selecting a place to publish your sensitive data, consider:

- Whether and how the repositories manage conditional access to data, and if so whether conditional access is managed by the repository or the data owner.

- Whether your institutional, employer, funder, or publisher mandates or recommends a particular repository

Other considerations of where to publish the data may include:

- Whether your research discipline has conventions around where to publish data
- Whether you want to publish your metadata and data in the same place
- What metadata the repository requires
- The format in which the repository requires your data
- Any financial costs incurred
- Whether the repository enables tracking of data citations by allocating the data a unique identifier (see the ARDC guidance on citation and identifiers)

Formatting data for publication

To maximise the benefits of publishing sensitive data, it is important that your data is formatted in a way that allows others to properly understand and re-use it. Particular care should be taken if the original data has been modified to manage sensitivity. The dataset you submit to a repository for publication should be[23]:

- Cleaned; checked for errors, outliers, duplicates, and missing data (and genuinely missing data should be annotated as such)
- Annotated; variables and objects (and their categories if relevant) should be clearly labeled and described and a key provided if necessary. If variables or objects are modified from their original form during de-identification, note this.
- In a format that is open, and easy to transform and archive. Strive to use a data format that is easy to read and easy to manipulate in a variety of commonly-used operating systems and programs. Non-proprietary ('open') formats are also recommended to enhance accessibility. Information about the data format should be provided in the metadata record. If a specialised program is required to read and analyse the data, then this should be provided alongside the data where possible.

Check your chosen data repository for any specific requirements about data formatting.

REUSING SENSITIVE DATA

Sharing sensitive data that you did not collect

What do you do about data publication and sharing if you are not the data owner? What if you are a data reuser or 'secondary data user'? This question is becoming increasingly common in response to requirements of journals to publish data alongside scholarly articles.

In most instances, you cannot publish data that you did not collect because you do not own the copyright for that data. Exceptions are data that have been licensed by the data owner to allow redistribution.

A note on research ethics: Like original researchers, secondary data users have an 'obligation to ensure that the data are used responsibly and respectfully, and that the privacy of participants is safeguarded'[1]. Thus, data re-users must uphold any conditions of data use that were specified to the participants by the original researchers, as well as those conditions of re-use outlined by the data owner or manager when access for re-use was granted.

Cite reused data

If you are a data re-user, you should reference or cite the original source of the data in all the articles, presentations, and grant applications based on that data. This enables the data owner to track the use of their data, as well as linking your work with other articles based on this dataset. See the ARDC guidance on data citation for detailed information on how to do this.

Recommendations:

- If the data you are re-using are licensed, follow the conditions of that licence regarding data sharing (typically termed 'redistribution'). If redistribution is allowed, you may be able to share the data and attribute to the data owner. Alternatively, cite the data source in any public description of your research so that others can also request access.
- If the data you are re-using is not licensed, contact the data owner or manager for instructions about publication and sharing.

Whenever you are reusing data, you should cite the original source in all scholarly outputs.

ACKNOWLEDGMENTS

ARDC acknowledges and thanks Dr. Greg Laughlin (Principal Policy Adviser, ANDS), Baden Appleyard (Barrister-at-Law, Queensland Bar), Professor Michael Martin (Research School of Finance, Actuarial Studies and Applied Statistics, The Australian National University; Chair, Humanities and Social Sciences Delegated Ethical Research Committee; Chair, Science and Medical Delegated Ethical Research Committee, The Australian National University), and Jeff Tranter (Environmental Resources Information Network; ERIN), for their invaluable advice and comments during the writing of this Guide.

MORE INFORMATION

Related ARDC Guidance

- Publishing and sharing sensitive data flowchart
- Data sharing considerations for Human Research Ethics Committees
- De-identification

Relevant National Guidelines

- National Health and Medical Research Council, Open Access Policy
- National Health and Medical Research Council, Australian Research Council, the Australian Vice-Chancellors' Committee, National Statement on Ethical Conduct in Human Research (2007, updated 2018).
- Australian Government Department of the Environment Sensitive Ecological Data – Access and Management Policy V1.0

Defining sensitive data

- Australian Government, Privacy Act 1988
- Office of Australian Information Commissioner, Australian Privacy Principles Quick Reference.
- Commission AGALR, The Privacy Act: Some important definitions
- Atlas of Living Australia, What is Sensitive Data?.

- Global Biodiversity Information Facility, Guide to Best Practices for Generalising Sensitive Species Occurrence Data.
- NSW Government Department of Planning, Industry and Environment, Sensitive Species Data Policy.

Preparing data for deposit

- UK Data Service 'Depositing Sharable Survey Data'.
- Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ. 2010;340:c181

Related international guides

- Digital Curation Centre (UK), How to Develop a Data Management and Sharing Plan.
- Inter-university Consortium for Political and Social Research (ICPSR), Guide to Social Science Data Preparation and Archiving.

REFERENCES

1. National Health and Medical Research Council, Australian Research Council, the Australian Vice-Chancellors' Committee. National Statement on Ethical Conduct in Human Research (2007, updated 2018). Canberra. Available from: https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018
2. National Health and Medical Research Council, Australian Research Council, Universities Australia. Australian Code for the Responsible Conduct of Research. Canberra: NMHRC. Available from: https://www.nhmrc.gov.au/research-policy/research-integrity/2018-australian-code-responsible-conduct-research.
3. National Health and Medical Research Council, Open Access Policy, Available from: https://www.nhmrc.gov.au/about-us/resources/open-access-policy
4. National Institutes of Health. NIH Data Sharing Policy. Available from: https://grants.nih.gov/grants/policy/data_sharing/

5. PLoS Medicine. Data Availability Policy. Available from:
http://journals.plos.org/plosmedicine/s/data-availability
6. Wellcome Trust. Policy on data management and sharing. Wellcome Trust. Available from:
http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm.
7. Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. http://dx.doi.org/10.1371/journal.pone.0000308
8. Loder E, Groves T. The BMJ requires data sharing on request for all trials. BMJ 2015;350:h2373 doi: https:/doi.org/10.1136/bmj.h2373 (Published 07 May 2015)
9. Australian Government: Privacy Act 1988. Available from:
https://www.legislation.gov.au/Details/C2020C00025
10. Office of Australian Information Commissioner. What is Personal Information? Australian Government. Available from:
https://www.oaic.gov.au/privacy/your-privacy-rights/your-personal-information/what-is-personal-information/..
11. Commission AGALR. Australian Privacy Law and Practice (ALRC Report 108). ALRC. Available from:
http://www.alrc.gov.au/publications/6.%20The%20Privacy%20Act%3A%20Some%20Important%20Definitions/sensitive-information
12. Office of the Australian Information Commissioner. Privacy in Your State. Australian Government. Available from: https://www.oaic.gov.au/privacy/privacy-in-your-state

13. Tann J, Flemons P. Our secrets are not your secrets: Proposed national policy and sensitive data report. Atlas of Living Australia. Available from:
http://www.ala.org.au/wp-content/uploads/2010/07/ALA-sensitive-data-report-and-proposed-policy-v1.1.pdf.
14. Australian Government Department of the Environment (2016) Sensitive Ecological Data Access and Management Policy. Available from:
http://www.environment.gov.au/system/files/resources/246e674a-feb1-4399-a678-be9f4b6a6800/files/sensitive-ecological-data-access-mgt-policy.pdf
15. Chapman AD, Grafton O. Guide to Best Practices for Generalising Sensitive Species Occurrence Data. Copenhagen: Global Biodiversity Information Facility. Available from:
https://www.gbif.org/document/80512/guide-to-best-practices-for-generalising-sensitive-species-occurrence-data.
16. Office of Australian Information Commissioner. Australian Privacy Principles guidelines: Privacy Act 1988. Canberra. Available from: https://www.oaic.gov.au/agencies-and-organisations/app-guidelines/
17. Iversen A, Liddell K, Fear N, Hotopf M, Wessely S. Consent, confidentiality, and the Data Protection Act. BMJ. 2006;332(7534):165-9.
18. McGuire AL, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, et al. To share or not to share: a randomized trial of consent for data sharing in genome research. Genetics in medicine : official journal of

the American College of Medical Genetics. 2011;13(11):948-55.

19. Inter-university Consortium for Political and Social Research. Recommended Informed Consent Language for Data Sharing. ICPSR. Available from: http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confientiality/conf-language.html .

20. Inter-university Consortium for Political and Social Research. Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle. Ann Arbor, MI: ICPSR. Available from:http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf.

21. UK Data Archive. Example consent form. UK Data Archive. Available from: http://www.data-archive.ac.uk/media/112638/ukdamodelconsent.pdf.

22.National Health and Medical Research Council. Guidelines approved under Section 95A of the Privacy Act 1988. Available from: https://www.nhmrc.gov.au/about-us/publications/guidelines-approved-under-section-95a-privacy-act-1988

23. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ. 2010;340:c181.