ANDS Guide

Publishing and sharing sensitive data

Level: Awareness

Last updated: 3 February 2017

Web link: **www.ands.org.au/guides/sensitivedata**

Key messages from the ANDS Guide to Publishing and sharing sensitive data

- The advantages of publishing your sensitive data will probably far outweigh any potential disadvantages when simple and appropriate steps are taken.

- Publishing your data, or just a description of your data (that is the metadata), means that others can discover it and cite it.

- You can publish a description of your data without making the data itself openly accessible.

- You can place conditions around access to published data.

- Sensitive data that has been de-identified can be shared.

Contents

1. Introduction

1.1   About this guide

"Despite journal and funder policies requiring data sharing, there has been little practical guidance on how data should be shared..." Hrynaszkiewicz et al. 2010[1]

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 Poster' (Section 1.2). 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. It can be done in most cases!

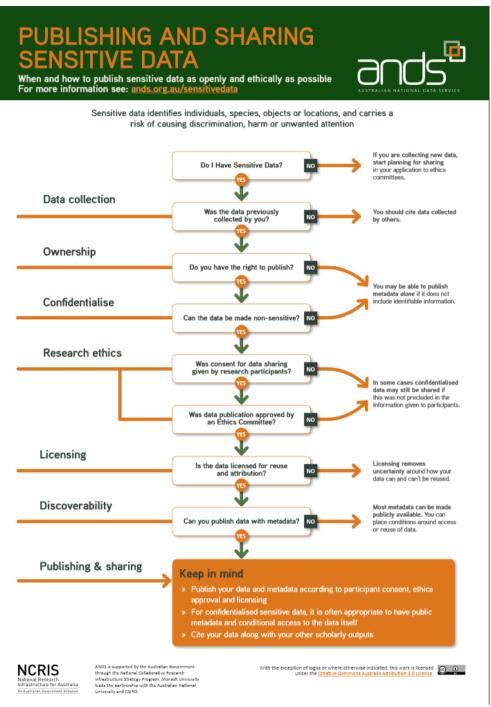**Introducing the ANDS Guide and a sensitive data sharing success story**

Dr. Sarah Olesen defines sensitive data and the steps involved in its publication. Associate Professor Leigh Tooth describes how the Longitudinal Study of Women's Health has safely shared health and social data to great success for 20 years. Watch on YouTube [38mins].

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 data repository.

1.2   Publishing and sharing sensitive data poster

http://www.ands.org.au/__data/assets/pdf_file/0010/385309/sensitive-decision-tree.pdf

1.3   Why publish and share sensitive data?

Box 1. Quick Definitions

Data publication versus sharing: Publication occurs when data are 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 are made available to others, but does not always accompany publication (e.g. when data are 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 institutional-specific, discipline-specific, and general repositories.

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

Sensitive data has often been excluded from discussions about data publication and sharing. It was believed that sharing sensitive data is not ethical or that it is too difficult to do safely. This opinion has changed with greater understanding and use of methods to 'de-sensitise' (i.e. de-identify) data; that is, modify the data to remove information so that participants or subjects are no longer identifiable, and the capacity to grant 'conditional access' to data. Requirements of publishers and funding bodies for researchers to publish and share their data have also seen sensitive data sharing increase.[2][3][4][5]

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.[6][7] And in most cases, this can be done by removing or modifying aspects of the dataset that make its subjects identifiable or attaching conditions to data access and reuse.

The advantages to sharing data go beyond meeting publisher or possible 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 [8]
- 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.[9]
- Storing your data in a public repository affords secure and ongoing storage that may not be available in your current or changing research environment.

2. What are sensitive data?

2.1   Defining sensitive

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.

2.2   Data about people

Sensitive human data most commonly refers to sensitive *personal information*. Personal information can be used to identify a person or group of people.[10] [11] [12] 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.

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

Specialised information about sharing data from or about Aboriginal and Torres Strait peoples and practices can be found in:

- AIATSIS Collection Access and Use Policy
- AIATSIS Guidelines for Ethical Research in Australian Indigenous Studies

2.3 Beyond human data

Sensitive data can also include data that reveals the location of rare, endangered or commercially-valuable species, or other conservation efforts.[13] The need for an agreed definition of sensitive environmental and biodiversity data is well documented.[13][14] The Australian Government Department of the Environment[15] 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. The lack of clarity surrounding the definition of sensitive ecological data is, at least in part, due to the sometimes-transitional nature of data sensitivity 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)[14] 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. Particularly when the publication or sharing of this data may increase that risk of harm and/or result in an adverse effect on the species. This GBIF report further recommends that data owners clarify which elements of the dataset trigger this sensitivity, and clearly document that decision in the metadata for the dataset.

2.4 An illustration of sensitive ecological data

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 (Section 5).

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. This latter aspect relates to ethical considerations about data publication (see Section 5).

2.5 Sensitive by context

Some data are born sensitive, some achieve sensitivity, and some have sensitivity thrust upon them!

What is considered to be sensitive may differ with time and across groups of people or subjects. Non-sensitive data may become sensitive by their 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[12]. 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 are not obviously sensitive (i.e., does not include names or dates of birth), or has been de-identified, 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** is 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 make 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 re-consider triangulation with the introduction of new data; that is, following data linkage.

3.   I have sensitive data - now what?

3.1   How to publish and share sensitive data

The previous section describes sensitive data in its original form. In the ANDS Guide to De-identification we discuss ways you can modify original data so that it is *no longer sensitive*. This kind of modified data is still highly valuable and re-usable to some other researchers without placing research participants or subjects at risk. And the data can be published and shared with fewer legal and ethical restraints.

In most instances, de-identified sensitive data is published with conditional access. This means that a description (i.e., 'metadata') is published in a data repository and its discoverability is not restricted - i.e., anyone can find and read a description of the data. And the data itself is accessible once some conditions (set by the researcher and/or

the repository) are met. The public description of the data includes this information about access and a link to how/where to apply for access.

3.2   What's legal?

Under the Australian Privacy Act 1988[10], 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. In other words, de-identified sensitive data can legally be shared. The *Australian Privacy Act 1988* 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). 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.

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). This activity is, however, explicitly condoned in the *Australian Privacy Principles* of the *Privacy Act 1988* 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 our section on Ethical Considerations (Section 5).

Key points on the legality of sharing sensitive data:

- Sensitive data cannot be published and shared in its original form in almost all cases.
- De-identified data is no longer sensitive and can be shared.
- The process of de-identifying human and personal data is allowed under the *Australian Privacy Act 1988.*[16]

---

4.   De-identifying data

The goal of de-identification for data sharing is 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; or, if modified to this extent would render 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 ANDS Guide to De-identification.

4.1   Ecological data

Methods for environmental and biodiversity data are less established than those for human data. For human data, sensitivity usually relates to the identification of individuals. Thus, identifying information must be removed for de-identification.

However, the sensitivity of ecological data typically does not typically 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 in Section 2.4 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 de-sensitised ecological data can be found in Table 1 of the Australian Government Department of the Environment's 'Sensitive Ecological Data – Access and Management Policy V1.0'[15].

4.2   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.

4.3   While you're at it; formatting data for publication

The dataset you submit to a repository for publication should be[1]:

- 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 labelled 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.

5.   Ethical considerations

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. The ethical management of data must be the primary concern of researchers to maintain participants' trust and research integrity.

Check: Are your data 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.

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

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 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.[6]

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.

5.2   Consent forms requesting data publication and sharing

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

1. Avoid 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 and possibility other conditions, such as approval by the original research team (see Section 7.1 on Conditional Access). 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.[6]

5.3   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 reusers.

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

Sensitive data can legally be shared without explicit consent from research participants if:[6]

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*, **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.

> **Recommendation**
> If you are unsure whether you meet these requirements, it is recommended that you seek the opinion of colleague/s outside your research project or your local Human Research Ethics Committee.

5.5   Sharing sensitive data that you did not collect

What do you do about data publication and sharing if you are not the data owner? That is, you are the data reuser or 'secondary data user'.

Data reuse is already relatively widespread in some research disciplines that deal in de-identified data, in fields such as epidemiology. The question above 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 (e.g., see AusGOAL's endorsed Creative Commons licences).

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*'[6]. Thus, data reusers must uphold any conditions of data use that were specified to the participants by the original researchers, as well as those conditions of reuse outlined by the data owner or manager when access for reuse was granted.

6.   Data citation

If you are a data reuser, you should also 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 ANDS Guide to Data Citation for detailed information on how to do this.

> **Recommentations:**
>
> •   *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.

7. Making data discoverable

A public metadata record in a repository is the best and easiest way to make your data discoverable by others. This is a description of your data in a public catalogue.

> You can publish a description of your data without making the data themselves 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. A public metadata record allows others to find and cite your data. If the data themselves are not published alongside the metadata record (most often via a link in the record, see example below), the record also provides the potential reuser with information about how the data can be accessed.

It is common for researchers with sensitive data to want to publish and share their data but still hold concerns about open publication or access. This is largely amenable by publishing the data with conditional access.

7.1   Conditional access to data: What is it; how do I do it?

Conditional access occurs when a metadata record is available to the public (i.e., published in a public repository) but access to the data themselves occurs only after pre-determined 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
- Agreed 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 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

Others can use the data for new and valuable purposes. Ideally, research data should be widely disseminated, whether or not the data creator or manager considers it valuable. It is impossible to know what applications of the data may exist in the future!

7.2 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 health." Professor Gita Mishra, ALSWH Director

Metadata record for the ALSWH

'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.

7.3 Licensing

All Australian data intended for reuse should have a licence. This includes your de-identified dataset. A licence is a document that clearly sets out how the data can be used and attributed to the original data owner. Without a licence, it is unclear how your data can be reused and this may discourage the potential reuser.

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).

AusGOAL provides a suite of endorsed licenses (ranging from few to more conditions of reuse) and a template for a Restrictive Licence. On the AusGOAL website you can select and follow simple steps to apply one of their licences to your data, or design a Restrictive Licence to suit your purposes. AusGOAL is endorsed by the Australian Governments.

It is important to note that applying an open licence to your data does not allow you to publish sensitive data, or act as a substitution for data de-identification. Sensitive data remains sensitive even with a licence, and thus cannot be published without participant consent. Accordingly, this Guide recommends that you apply a licence, publish, and share your data *after* it has been de-identified as described above, or you have consent to share identifiable data and it is being made available through conditional access arrangements.

For more detail see Licensing for data reuse and the ANDS Guide to AusGOAL.

8. Your right to publish

In most cases, the person or institution that publishes the data must also hold the appropriate rights (such as copyright) to do so. An AusGOAL endorsed licence can only be applied by to the rights holder of that data.

8.1 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 ANDS Guide to Copyright, Data and Licensing.

8.2 Where should I publish my 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 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 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 your 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 your data a unique identifier (see DOIs)

9. Acknowledgments

10. More information

Related ANDS Guides

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

Relevant National Guidelines

- National Health and Medical Research Council, NHMRC Statement of Data Sharing
- National Health and Medical Research Council, Australian Research Council, the Australian Vice-Chancellors' Committee, National Statement on Ethical Conduct in Human Research (2007, updated May 2015).
- 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, Privacy fact sheet 17: Australian Privacy Principles.
- Commission AGALR, The Privacy Act: Some important definitions
- Atlas of Living Australia, Data sensitivity.
- Global Biodiversity Information Facility, Guide to Best Practices for Generalising Sensitive Species Occurrence Data.
- NSW Government Department of Environment, Climate Change and Water, 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

- UK Data Service 'Depositing Sharable Survey Data'.

- 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.

11. References

1. 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.

2. National Health and Medical Research Council. NHMRC Statement of Data Sharing (2016). Available from: https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing?

3. National Institutes of Health. NIH Data Sharing Policy. Available from: https://grants.nih.gov/grants/policy/data_sharing/

4. PLoS Medicine. Data Availability Policy. Available from: http://journals.plos.org/plosmedicine/s/data-availability

5. 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.

6. National Health and Medical Research Council, Australian Research Council, the Australian Vice-Chancellors' Committee. National Statement on Ethical Conduct in Human Research (2007, updated May 2015). Canberra. Available from: https://www.nhmrc.gov.au/guidelines-publications/e72

7. 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/guidelines/publications/r39.

8. 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

9. 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)

10 . Australian Government: Privacy Act 1988. Available from: https://www.legislation.gov.au/Details/C2016C00979

11. Office of Australian Information Commissioner. Privacy fact sheet 17: Australian Privacy Principles. Australian Government. Available from: http://www.oaic.gov.au/privacy/privacy-resources/privacy-fact-sheets/other/privacy-fact-sheet-17-australian-privacy-principles..

12. . 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

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. Chapman AD, Grafton O. Guide to Best Practices for Generalising Sensitive Species Occurrence Data. Copenhagen: Global Biodiversity Information Facility. Available from: http://www.gbif.org/resources/2760.

15. 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

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.

Feedback?

We welcome your feedback on this guide. Please email contact@ands.org.au with any comments or questions.

About ANDS

The Australian National Data Service (ANDS) makes Australia's research data assets more valuable for researchers, research institutions and the nation.

ANDS is a partnership led by Monash University in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).