Australian Government

Department of the Prime Minister and Cabinet

Best Practice Guide to Applying Data Sharing Principles

Version: 15 March 2019

Contents

Glossary of terms

Data Custodian: The agency that collects or generates data for any purpose, and is accountable and responsible for the governance of that data.

Data Protections: Changes made to data to minimise the likelihood of identifying the Data Provider.

Data Provider: An individual, household, business or other entity that supplies data, or has data about them supplied by a third party, to a government agency.

Data Release: Making data publicly available with no or few restrictions on who may access the data and what they may do with it.

Data Sharing: Making data available to another agency, organisation or person under agreed conditions.

Data Sharing Agreement: A formal arrangement between a data custodian and another agency, organisation or individual that details conditions under which data is shared and used.

Disclosure Risk: The combination of likelihood and consequence that information about an individual, organisation or other entity is revealed or provided to an unauthorised person or entity.

Direct Identifier: Information which, by itself, is able to identify an individual, organisation or other entity. Examples of direct identifiers are name, latitude/longitude, driver's licence number and Australian Business Number.

Particularly Sensitive Data: Any data where unauthorised disclosure would likely lead to adverse consequences for the individual, agency, organisation or Australia in general. Data which is of a personal, legal, commercial, security or environmental nature may be considered particularly sensitive. This is broader than the *Privacy Act 1988* definition of sensitive data which is defined as a subset of personal information and limits how it can be collected and used.

Personal Information: Information or an opinion about an identified individual, or an individual who is reasonably identifiable:
(a) whether the information or opinion is true or not true; and
(b) whether the information or opinion is recorded in a material form or not.[1]

Responsible Officer: A senior person in an organisation who has the legal authority to agree to conditions of shared data use on behalf of that organisation.

[1] *Privacy Act 1988*

Introduction

Context

The Australian Government holds vast amounts of public sector data that it collects from individuals and businesses, or generates through administrative functions of government agencies. This data has significant potential to inform policy development, evaluate programs, contribute to economic growth, and support innovation, for the benefit of all Australians.

Acknowledging the value of public sector data, and the need use it efficiently and with appropriate safeguards, the Australian Government established the Office of the National Data Commissioner (ONDC) in July 2018. The ONDC is responsible for implementing a data sharing framework that improves access to and re-use of public sector data, while maintaining data privacy and security. In this context, *data sharing* is the provision of access to data in a controlled manner. *Data release* means providing open access to data, i.e. making it publicly available for anyone to use.

The potential of public sector data can be realised in a number of ways. Data sharing allows re-use of existing data to deliver public benefit and the creation of new datasets to provide rich insights about communities, families, industry, the environment and the economy. However, data sharing must be managed carefully and safely to ensure the public trusts how Australian Government agencies handle the data they hold.

About this Guide

This Guide has been written to assist agencies holding Australian Government data (*data custodians*) to safely and effectively share the data they are responsible for by using five *Data Sharing Principles* (the Principles).

Where there is a clear public benefit, data custodians may seek to share data in a controlled manner with a range of users, such as Government agencies, the academic research community and, in some cases, the private sector.

This Guide has been structured to assist data custodians to consider the appropriate safeguards to apply before sharing Government data, and to promote more flexible, principle-driven data access arrangements.

Part 1 contains information and questions for data custodians to consider prior to sharing data, such as the data sharing maturity of an organisation and their approach to managing risk. Part 2 explains each of the Principles in a practical order, beginning with the project assessment. It provides examples of how each principle operates and poses questions to help data custodians apply them. Part 3 includes further guidance on how to manage data sharing agreements once they are in place.

The aim of this Guide is to provide an introduction to the Principles. The Principles are designed to enable safe and appropriate data sharing. A data custodian will need to be flexible in applying the Principles by taking into account the context in which the agency intends to share data, and may need to consider other questions than those in this Guide.

This Guide will be published at www.datacommissioner.gov.au and updated periodically.

The Data Sharing Principles

There is a growing imperative for public sector data to be used more effectively to improve government service delivery and solve complex policy issues that can't be addressed when data remains in siloes across government. However, for many data custodians, there may be barriers to sharing data easily. For example, there may be some concerns about sharing an agency's data and exposing it to external scrutiny which may lead to a decision not to share data, or to apply unnecessary protections to the data, and may significantly reduce its usefulness. While these are important concerns, with appropriate risk management, they can be weighed against the potential benefits to the public that can arise from data sharing.

The ONDC, together with the Australian Bureau of Statistics (ABS), has developed the Principles to support agencies' 'responsibility to share' by providing an approach for effectively managing the risks associated with data sharing. Applying the Principles can enable safe and effective sharing of data held by the public sector in a way that delivers public benefit, protects privacy and maintains confidentiality.

The Data Sharing Principles are based on the Five Safes Framework. Originally developed in the United Kingdom at the Office of National Statistics, the Five Safes Framework is an internationally recognised approach to disclosure risk management. The ONDC has adapted the Framework as a set of principles to emphasise the broad set of considerations related to data sharing in Australia.

The Principles enable a privacy-by-design approach to data sharing by balancing the benefits of using government data with a range of risk-management controls and treatments (particularly those managing disclosure risks). By focusing on controls and benefits, instead of merely reducing the level of detail in the data to be shared, the Principles can assist with maximising the usefulness of the data.

For example, an agency may be unwilling to share a dataset publicly because of the risk of identifying the individuals who provided the data. However, the same agency may be comfortable with sharing that dataset with only basic data protections in place, such as the removal of names and addresses, as long as it is only accessed by authorised researchers in a secure environment. Alternatively, an aggregated form of the same data which does not identify any individual person or entity may be made available on a website for public use. This flexible approach may increase access opportunities and could lead to improved outcomes for research and decision-making, while still ensuring appropriate safeguarding of the data.

The Data Sharing Principles are:

1. Projects: Data is shared for an appropriate purpose that delivers a public benefit.
2. People: The user has the appropriate authority to access the data.
3. Settings: The environment in which the data is shared minimises the risk of unauthorised use or disclosure.
4. Data: Appropriate and proportionate protections are applied to the data.
5. Output: The output from the data sharing arrangement is appropriately safeguarded before any further sharing or release.

Not all data sharing arrangements will require extensive consideration of each Principle. For example, it may only be necessary to consider the data principle in circumstances where it is intended that the data will be published or accessible in a publicly available space, such as on data.gov.au.

Part 1 – Before applying the Data Sharing Principles

The data sharing request

Data sharing may be initiated by a data custodian; however it is generally initiated by a request to a data custodian for data. Requests may come from another government agency, the private sector or the research sector. These requests can be managed more easily if the agency has a catalogue of data available to enable users to discover data more efficiently.

A data request is used to initiate the consideration of a data sharing project; the contents may also support the development of a formal Data Sharing Agreement which clearly articulates the arrangements, terms and conditions of a data sharing project. The request should also describe the purpose for which the data will be used, and can be used to make an initial assessment of the appropriateness of a project. Each data request is likely to include many if not all of the following requirements:

- demonstrate an appropriate aim, in line with a relevant purpose test if applicable;
- demonstrate a public benefit;
- show that legal, ethical and moral considerations have been addressed;
- state what data will be used and why it's required;
- state the timeframes for which the user needs the data, and the expected outputs and outcomes;
- state who (either named individuals or groups) will be working on the project; and
- demonstrate feasibility (i.e. the data is suitable/appropriate for answering the question).

Is the data available and suitable?

It is essential for a data custodian to assess the request and identify the primary source(s) of data that could be shared in order to satisfy the request. Initially this may involve a data custodian ensuring that its available data is broadly relevant to the request. A data custodian will have the best understanding of what can and cannot be achieved with the data they hold. Discussion between the data custodian and the requestor can explore how a request may be supported, including whether other agencies may need to be involved. It is also important to begin considering the project's benefit to the public - this is addressed in more detail in Part 2 of this Guide.

Can the data be shared legally?

The data custodian will need to confirm that there is a legal basis to share the data. Some laws prohibit certain people or organisations (for example, those who are not Australian citizens) from accessing Australian government data. Data custodians need to be aware of these restrictions and communicate them clearly to prospective users. Data custodians should explore how they can share data legally rather than simply dismissing a request to access data due to perceived legislative restrictions.

Is there any particular sensitivity in the data?

Data custodians will need to determine whether, or to what degree, the data is considered particularly sensitive – for example personal, commercial, environmental, national security or legal sensitivities may be evident in the data. It is also important to consider how the sensitivity of data may change following the application of the Data Sharing Principles. For example, endangered species data may include detailed information about the location of habitats if access to the data is limited to authorised users, but this same information may need to be removed if it were to be released publicly.

Data Sharing Agreements

Data sharing agreements are made between a data custodian and the organisation receiving their dataset (for example, other government agency, research institution, non-government organisation, private company etc.). These agreements may include how a purpose test is satisfied and details of projects covered by the agreement. It should also specify what the data can and can't be used for, and provide information on any sanction that may be imposed if the terms and conditions of the agreement are not adhered to (this may include reference to legally enforceable sanctions available under any relevant law).

In the data sharing agreement, the responsible officer of the organisation receiving or accessing the data would agree that all users within their organisation will abide by the terms and conditions for accessing the data. The responsible officer may be required to provide and maintain a list of individuals (or groups of individuals within an organisation) that are accessing data under the agreement. In some cases, individual users within an organisation may also need to agree to conditions of use, which may be part of authorisation criteria.

It is best practice to make data sharing agreements publicly available to maximise transparency.

Consider how to best meet the user's needs

It is important to consider the specific needs of the requesting person or organisation when determining how to support a data sharing arrangement. For example, data custodians are often aware of what researchers and research organisations are interested in and should be proactive in publicly releasing data where possible. Researchers are able to access publicly available data to gain early insights, which in turn assists them in targeting their data requests.

Once a request is received, an agency, in consultation with the requester, will need to determine the most appropriate sharing arrangement. Options for sharing data include:

a) the data itself is shared – that is, the data is given to the user for them to work on within their own environment; or
b) access to the data is provided – that is, the data remains in the data custodian's environment and the user is granted some form of access to that environment (onsite, offsite or via an analysis service).

Data custodians have a responsibility to ensure that, where it is possible to do so legally and safely, data is shared in a way that serves the Australian public. This means, it is important to take a user-centred approach to enable the user, and the public, to get the most benefit from the data sharing. Further guidance on options for sharing data is included in Part 2 of this Guide.

Capability and culture

For some data custodians, sharing data may be a daunting prospect. Before sharing, it may be necessary to assess the internal skills and capability available, and seek additional data expertise or capabilities where necessary to effectively manage data sharing arrangements. There may also be internal cultural resistance, requiring data custodians to move from a culture of risk aversion to a culture of managing the risks associated with data sharing. The Principles and this document are designed to support this cultural shift.

Ensuring clear responsibility for each shared dataset

Some data sharing agreements may be between data custodians and an agency that is able to provide data services (for example, the Australian Institute of Health and Welfare manages many of the health-related data linking requests for the Commonwealth and States and Territories).

In this situation, there is joint responsibility, and the data custodian still retains accountability for appropriate use of the data under a data sharing agreement.

Governance and the Data Sharing Agreement

Developing appropriate governance for data sharing is a way of providing confidence for the data custodian, the Government and the public. A Data Sharing Agreement is the means of ensuring all aspects of the data sharing, the participants and their responsibilities are documented.

Good governance requires transparent decision-making (for example, a record of the assessment of risks involved with the project) and this transparency can offer a constructive basis for engagement with the public. It is also a good idea for data custodians to set up streamlined processes to handle subsequent requests more efficiently.

Costs

The costs of data provision and access need to be considered. If costs are to be passed on to users, this should be communicated to the data users and documented in the data sharing agreement.

The ideas raised above are summarised in the list of questions below. It should be noted that this is not an exhaustive list. Advice from organisations that have experience in applying the Principles will be invaluable for data custodians who do not regularly share their data.

Questions to ask: Before applying the Data Sharing Principles

1. What is the benefit or value to individuals and society of sharing this data?
2. How can this data sharing arrangement be done legally?
3. Does the organisation have the data maturity to manage sharing (e.g. are the required data skills and capability available?). If not, is an external service provider needed to assist?
4. Has a data source been identified?
5. Is the data source fit for purpose (i.e. will it meet the user's needs?)
6. Is an external service provider needed to provide expert data services, (e.g. data linking, storage)?
7. Has the sensitivity of the data been assessed?
8. What costs are associated with preparing and sharing the data?
9. Have good governance processes been established for data sharing?
10. Are arrangements as streamlined as possible?

Part 2: Applying the Data Sharing Principles

Managing data sharing risks

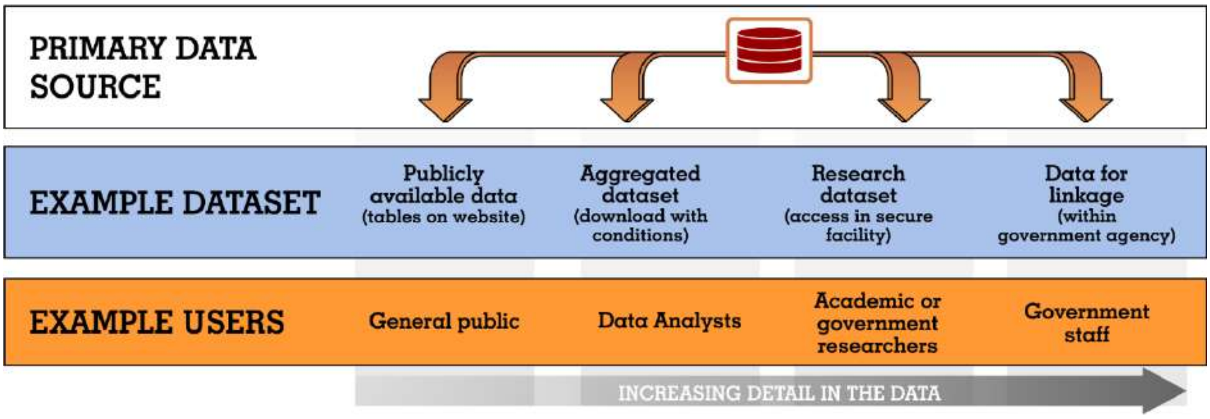
In order to encourage the safe sharing of data, the five Principles provide a disclosure risk management framework, which balances risks against public benefit. Each of the Principles can be considered as an adjustable control mechanism (for example, proportionately higher or lower levels of control on the environment in which the data is accessed). While each Principle can be considered independently, all five Principles need to be considered jointly to evaluate whether a particular instance of data sharing is a safe arrangement. It is the application of all the Principles together that can deliver a safe data sharing arrangement. Where application of the Principles cannot provide a safe data sharing arrangement, then the data custodian should not share that data.

Controls should be based on a realistic assessment of the likelihood and consequence of a risk occurring and be made in the context of organisational risk tolerance, rather than based on hypothetical worst case scenarios.

Adjustable controls provide flexibility to potentially share the same primary data source in multiple ways in order to service users with different needs. Throughout this Guide an example of a primary data source containing detailed information such as income, employment and location will be used. From this source other datasets (see Diagram 1) could be created:

- *Publicly available data* such as a limited set of tables (for example, household income by location and personal income by profession) made publicly available on a website. This is actually a data release (i.e. some form of the data is available to anyone) rather than sharing, but is included for completeness.
- *Aggregated dataset* such as complex tables or aggregated records, could be downloaded for basic research by users with a greater understanding of the data.
- *Research dataset* with direct identifiers removed could be accessed by authorised researchers in secure facilities.
- *Integrated dataset* where the primary data could be linked to a dataset from another agency (for example, linking to household income data to school attendance data) to enhance its use for other, carefully controlled, projects.

Diagram 1: data from a single data source can be shared in many different ways


PRIMARY DATA SOURCE
EXAMPLE DATASET
Publicly available data (tables on website)
Aggregated dataset (download with conditions)
Research dataset (access in secure facility)
Data for linkage (within government agency)
EXAMPLE USERS
General public
Data Analysts
Academic or government researchers
Government staff
INCREASING DETAIL IN THE DATA

Each data sharing scenario will require different levels of control under each of the Principles, but each instance of data sharing should be designed to provide an acceptable solution for the user's needs. A user may access data under more than one scenario as their requirements change, with each access arrangement assessed and managed through the application of the Principles.

1. Project Principle: Data is shared for an appropriate purpose that delivers a public benefit

The Project Principle addresses the intended purpose or use of the data in the data request. A data custodian needs to ask: "Is this use of the data appropriate?" The decision will be based around ethical, legal and public benefit considerations. Each data custodian is likely to have a different set of considerations, because each will operate in a different context.

Data sharing purpose test

Many government agencies will have a policy or legal requirement that data sharing may only be undertaken if the data satisfies a purpose test; for example, if the purpose is to inform:

- Government policy
- Research and development with a public benefit
- Program design, implementation, and evaluation, or
- Delivery of government services.

Assessment of data sharing projects

Each data sharing project (whether part of a broader Data Sharing Agreement or not) will usually require assessment which should be managed through a formal governance process. This may need to be established, or an existing one modified, to assess data sharing projects. Strong governance arrangements ensure that assessments are consistently applied, based on qualified opinions and that decisions are transparent. If an agency is new to data sharing, it may be necessary for a governance body to scrutinise all project proposals. As experience is gained, streamlining assessments may be desirable, so that project proposals are considered more efficiently (for example, by a small team, with only unusual or higher risk project proposals being considered by the governance body). This streamlining will allow for faster turnaround of project proposals, while also allowing for greater scrutiny where necessary.

To assist with the assessment, a data custodian can request some key aspects be included in a project proposal such as specifying requirements for ethics approval or consent from the original data provider. For example, being able to show that a project has been considered by a formal ethics committee approval process will demonstrate to both the requester, as well as the data custodian, that the project has no significant ethical barriers. Similarly, if informed consent is available from the data provider(s), this may reduce data custodian concerns.

There will be other considerations that may affect the project assessment process, such as costs of sharing and how the sharing may affect the organisation. For example, research that examines an agency's methodology might be perceived as a risk by the agency, but could equally be used as an opportunity for the agency to improve their methods. In this case, collaboration may be more beneficial than not proceeding.

The Project Principle allows agencies to consider whether sharing the data they are responsible for in a particular form would provide a public benefit. The four remaining principles address how to balance privacy, disclosure and other obligations and how to minimise risks.

An iterative process

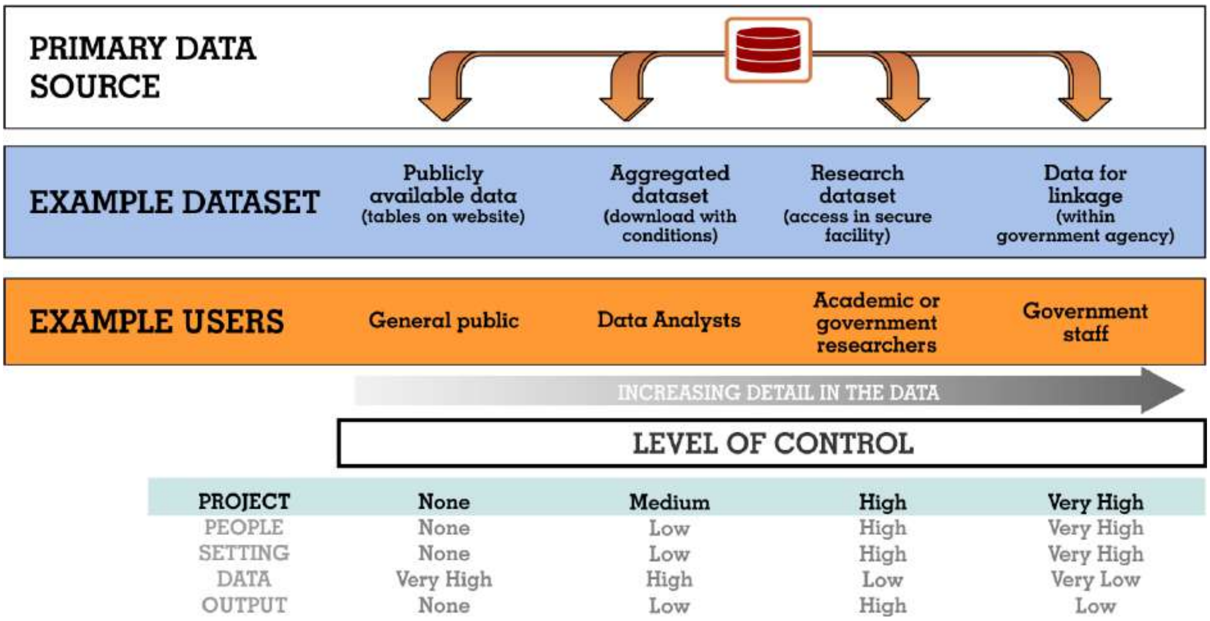
Project assessment and approval is the first step towards sharing data. The remaining four Principles will need to be considered independently and jointly, so that the project and the data sharing arrangement are executed in a safe manner. If consideration of the remaining Principles identifies risks that are unable to be addressed then the proposed project may need to be modified and the remaining Principles considered in the context of the changes.

Questions to ask: Project Principle

1. Is the project in the public interest and does it satisfy a purpose test?
2. Has all relevant information been provided to support assessment of the project proposal (e.g. who will access the data, for what purpose, over what period of time and what will happen to the data when the project ends)?
3. What processes or governance arrangements are needed to assess, monitor and oversee the project?
4. Who will make the assessment of whether to proceed with the project and do they possess the right capabilities to make the assessment?
5. Are there any restrictions (e.g. legal or data custodian imposed restrictions) on how the shared data may be used?
6. How will communication with applicants before and during the assessment of the project proposal be managed to maximise the likelihood of approval? What feedback will be provided?
7. Does there need to be ethics approval from a governance body that considers the ethics of the proposal?
8. Is consent from the original data providers required?
9. What collaboration opportunities could the project provide to improve organisational processes?

Applying the Project Principle: The diagram below shows that different levels of control need to be placed around the project depending on the use. While sharing of publicly available data does not require project controls to be applied (because the use of the data cannot be controlled by the data custodian), additional and stronger controls (such as limitations on the use and on-disclosure of the shared data) may be necessary as the detail in the data increases.

Diagram 2: the level of project control will depend on the level of detail being shared



PRIMARY DATA SOURCE

EXAMPLE DATASET	Publicly available data (tables on website)	Aggregated dataset (download with conditions)	Research dataset (access in secure facility)	Data for linkage (within government agency)
EXAMPLE USERS	General public	Data Analysts	Academic or government researchers	Government staff

INCREASING DETAIL IN THE DATA

LEVEL OF CONTROL

PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	None	Low	High	Very High
DATA	Very High	High	Low	Very Low
OUTPUT	None	Low	High	Low

2. People Principle: The user has the appropriate authority to access the data

Under the People Principle, the user may be required to undergo an authorisation process to assess the user's knowledge, skills and motivations to determine whether they can use (and in some cases store) any shared data appropriately.

Authorising users

The criteria for authorising users may have a legal basis (for example, a law may authorise a particular user to access data), or may serve to satisfy a data custodian that a user understands expectations when accessing shared data. In some cases, a data custodian may be able to use all or part of an authorisation process that has been developed by another agency in order to limit duplication.

Users may be authorised to access shared data for a particular project, or obtain a more general authorisation to access data for multiple projects. A more general authorisation could include the right to access data on an ongoing basis (for example, access to a dataset that is periodically updated by a data custodian). The data custodian will need to consider the scope of the authorisation in the context of each access request.

The following criteria should be considered by data custodians when authorising data users to access shared data. Not all criteria may be necessary (depending on the sensitivity of the data to be shared):

- A formal application by a user to become an authorised user.
- The user is part of an organisation that has an overarching agreement with a data custodian.
- The user provides evidence of technical ability in data analysis.
- The user has signed an agreement or legally binding undertaking which governs the access and use of the data to be shared.
- The user acknowledges their understanding of the sanctions or penalties that apply for a breach of an undertaking or agreement.
- The user is provided with training regarding their rights and responsibilities.

In some cases (for example, for especially sensitive data), other criteria may be desirable, such as:

- The user is required to hold a current Security Clearance at an appropriate level.
- Training is provided on specific technical aspects of working with the data.
- The user is able to demonstrate experience in using particularly sensitive data appropriately.
- The user is required to be seconded or employed by the organisation that holds the data.
- The user is required to be formally endorsed by a senior member of their organisation.

Training of users

International and Australian experience in data sharing has shown that the main cause of data breaches is people making mistakes when using data rather than failures of technology or deliberate misuse. For example, a user who has been given individual access to a secure dataset assumes they can share their access with a colleague who is not authorised to access the same dataset.

An effective approach to minimise mistakes is to provide training, ideally face-to-face. Training can provide benefits to all parties, ensuring everyone understands their obligations and responsibilities in a data sharing community. The training should emphasise the positive behaviours and attitudes necessary to use the data in a manner consistent with the requirements of the data sharing agreements. The legal consequences of data misuse need to be raised in training, but it is also important that users understand other non-legal penalties can apply, such as withdrawal of access to data, which may detrimentally affect all data users within a community. Training should also be simple, user-centred, positive and interactive (for example, data misuse scenarios can be used to highlight, and facilitate discussion about, legal, moral and procedural issues around data sharing).

The data custodian, or a training provider engaged by the data custodian, may deliver tailored training. It can be used to clearly convey that an organisation must meet its obligations, and that individual users must understand how to appropriately access, use and destroy data, consistent with the data sharing agreement.

An alternative to training may be to provide users with a “do and don’t” document. This approach should be used cautiously as it has been found to be less effective than face-to-face training, although it may be acceptable if controls in the other Principles are enhanced. Users are less likely to read a long document and it may not be able to effectively articulate all nuances of appropriate and inappropriate data use.

As a complement to the training, the data custodian may choose to periodically test users to ensure they understand their responsibilities, and can demonstrate appropriate attitudes and behaviours regarding safe data use. For example, providing a scenario relating to an observed procedural breach and requesting the user describe how they would respond to the event, rather than simply asking whether they should report an observed breach.

It is recommended that at a minimum, testing be conducted at the same time, or as soon as possible after, the training. Combining tools may also be appropriate in that a test could be used to train people by asking for their views on do’s and don’ts.

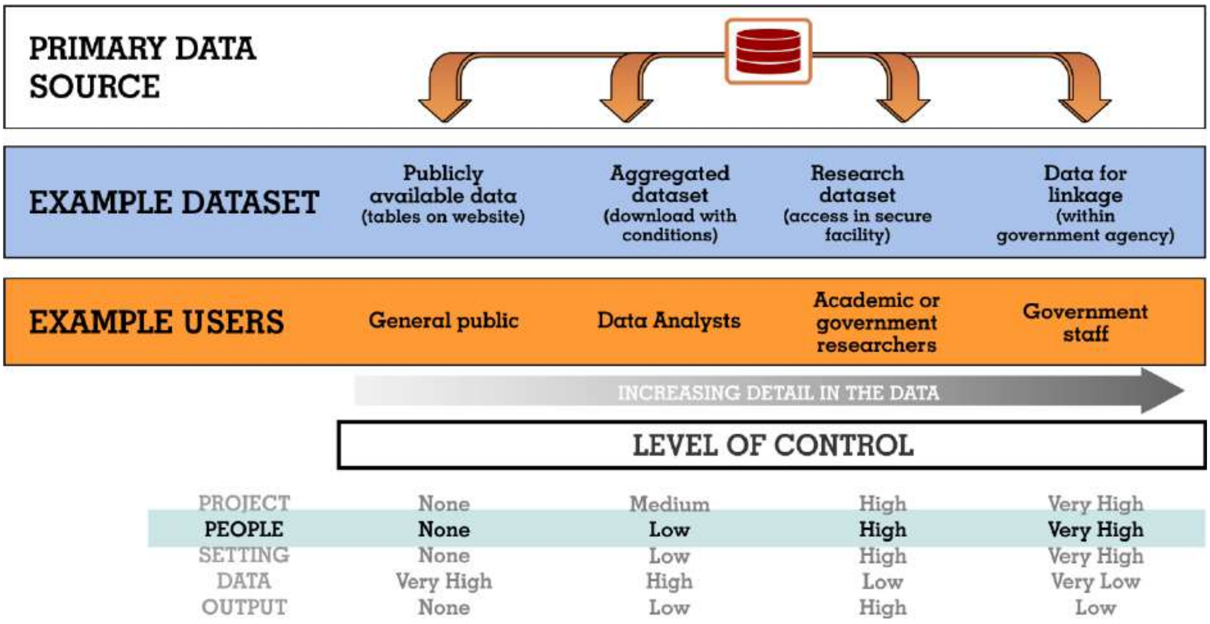
Questions to ask: People Principle

1. What, if any, process for authorisation is required for people or organisations to access data? Who facilitates this process? How long will any authorisation be valid?
2. Will a legally binding undertaking or agreement to govern the access and use be required? Who needs to complete this undertaking or agreement?
3. Does the user need to meet any specific criteria to access data (e.g. hold a current security clearance)? What are these criteria?
4. Does the user have a history of good data handling practices? Does the user need to seek endorsement of their data handling practices from a responsible officer of their organisation?
5. Does the user need to be trained in safe use, data storage and technical skills? Who develops and/or provides training?
6. What sanctions (legal and non-legal) need to be available for misuse of data? Are these clear to the user?
7. Are there any restrictions on who may apply to access the data (e.g. must be an Australian citizen, current affiliation with a particular research institution)?

Applying the People Principle: Having applied controls under the Project Principle, the diagram below shows that different levels of control will also need to be placed around the people — again depending on the user. While publicly available data does not require people controls to be applied (because the data custodian cannot control who accesses the data), additional controls are invariably necessary as the detail in the data increases.

For example, data analysts accessing an aggregated dataset might need instructions around the limitations of the dataset before they use it, researchers using a secure facility might need to undergo an authorisation process and appropriate training, or in circumstances where data is used for linking projects, users might be required to have an appropriate security clearance and be working in an appropriately secure data environment.

Diagram 3: the level of user authorisation and training will depend on the level of detail being shared



PRIMARY DATA SOURCE
EXAMPLE DATASET
Publicly available data (tables on website)
Aggregated dataset (download with conditions)
Research dataset (access in secure facility)
Data for linkage (within government agency)
EXAMPLE USERS
General public
Data Analysts
Academic or government researchers
Government staff
INCREASING DETAIL IN THE DATA

LEVEL OF CONTROL

PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	None	Low	High	Very High
DATA	Very High	High	Low	Very Low
OUTPUT	None	Low	High	Low

3. Settings Principle: The environment in which the data is shared minimises the risk of unauthorised use or disclosure

The Settings Principle considers whether all parties have taken reasonable steps to ensure data will be used in an appropriately safe and secure environment, i.e. one that minimises unauthorised use, access or loss of data. Data custodians need to consider the practical controls, in both physical and IT environments that can be put in place to control how data is stored, transferred and accessed. Controls will depend on whether the user will be provided access to the data (for example, via a web service, secure facility or API) or be given the data itself (for example, by download or physical media).

Physical environment

Physical controls may include:

- Working in an agreed location (for example, in a personal office, an access controlled room or at home).
- Direct and active supervision.
- Access during certain restricted times.
- Keeping physical copies (for example, CD, USB stick or printed) of data locked away and secured during transfer to a user.
- Making users aware of surroundings and taking care in open plan offices to avoid data being viewed on screen by unauthorised people.

IT environment

IT controls may include:

- IT security controls as mandated by the data custodian and/or Australian Signals Directorate.
- Data held in a subsystem within the data custodian's IT system.
- Closed IT environment with no external email or internet access.
- Role-based access.
- Two-factor authentication.
- Recording of access sessions, with auditing/review conducted in a transparent manner.
- Retention of data on a secure server or specific computer/drive with appropriate password and access protections.
- Requirements to provide evidence of data destruction at the end of project or project approval period.

An important aspect of this principle relates to training (often as part of the authorisation of users). This is done in order to help the user avoid mistakes and to satisfy the data custodian that the user can be reasonably expected to use and store the data appropriately. Training can reinforce responsibilities such as:

- Protecting work areas from oversight of unauthorised people.
- Maintaining a 'clear screen' (i.e. securing work stations appropriately when away).
- Not on-sharing information to unauthorised people.
- Requiring outputs that are intended for wider sharing to be approved by the data custodian.
- Reporting any security incidents to the data custodian as soon as practicable.
- Not removing any of the information from the approved setting without authorisation.
- Not sharing login details.

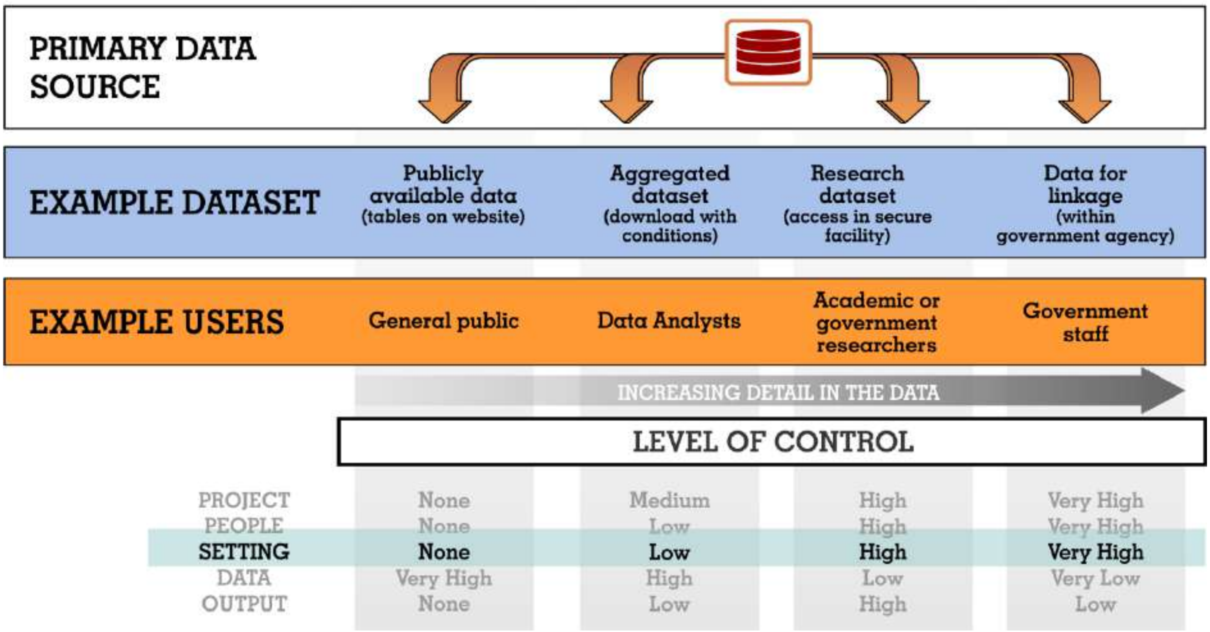
Questions to ask: Settings Principle

1. From what physical location(s) will the data be accessed?
2. Does there need to be auditing/checks of these locations?
3. What physical supervision is appropriate?
4. What IT security needs to be in place? Will the security classification of the data influence IT security requirements?
5. What electronic supervision as well as auditing/recording of use is available?
6. Is certification of physical and/or IT environment necessary? If so, by whom?
7. Do the controls limit misuse (by mistake and deliberate), interference, unauthorised access, modification, loss or disclosure?
8. Do users understand how to access the data safely in the IT and/or physical environment? Is training required?
9. How will data transfer into and out of a secure environment be managed?

Applying the Settings Principle: In considering what controls are necessary around the Settings, the following diagram shows that different levels will again be required. In a similar pattern to the Project and People Principles, while publicly available data does not require setting controls to be applied (because how the data is accessed cannot be controlled by the data custodian), additional settings are necessary as the detail in the data increases.

An example of controlling the environment is providing data to an academic researcher in the form of a downloaded file on a secure drive rather than transferring it to portable media. There are also closed IT systems with password-protected, role-based access available such as the ABS DataLab, the Secure Unified Research Environment (SURE) or the E-Research Institutional Cloud Architecture (ERICA) through which many government agencies undertake their work.

Diagram 4: the level of controls applied to the data environment will depend on the level of detail being shared



PRIMARY DATA SOURCE

EXAMPLE DATASET	Publicly available data (tables on website)	Aggregated dataset (download with conditions)	Research dataset (access in secure facility)	Data for linkage (within government agency)
EXAMPLE USERS	General public	Data Analysts	Academic or government researchers	Government staff

INCREASING DETAIL IN THE DATA

LEVEL OF CONTROL

PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	**None**	**Low**	**High**	**Very High**
DATA	Very High	High	Low	Very Low
OUTPUT	None	Low	High	Low

4. Data Principle: Appropriate and proportionate protections are applied to the data

The Data Principle focusses on what treatment of the data (for example, data minimisation, aggregation, removing direct identifiers, or suppressing individual records) is necessary to control for risks that could not be addressed by the Project, People and Settings Principles.

Limitations of the Data Principle

It is important to understand the difference between the Data Principle and the Output Principle. The Data Principle applies controls (such as removing direct identifiers and other confidentiality treatments) to the whole dataset available to the user, whereas the Output Principle applies controls to results that are to be made public or available for further sharing by the authorised user. The Data Principle protects data going from the data custodian to the data user. The Output Principle protects the data subsequent to leaving the data user.

Retaining a user-centred perspective as outlined in Part 1 means that restrictions should not be applied to the data unless there is good reason to do so. Every restriction applied to a dataset may reduce its potential usefulness. An appropriately authorised user might have access to the highest detail of data possible, in a controlled environment, for an approved purpose. In datasets where confidentiality may be a concern, most of the analytical outputs created by users will protect the data to some degree (for example, produce a table, regression, model, summary, etc.). In these cases it may only be necessary to remove the direct identifiers since confidentiality and privacy concerns are able to be controlled by the application of the Project, People, Settings and Output Principles.

Data custodians should keep in mind that the Output Principle is there to control for any residual risk in making results public or available for further sharing (this is addressed in more detail under the Output Principle section of this Guide).

Further data restrictions may be appropriate if other controls are insufficient to manage risks. The level of control applied to the data will depend on the sensitivities associated with the data.

As a minimum, this Guide recommends that removal of direct identifiers is applied in most cases of data sharing. Identifiers should only be retained if they are absolutely critical for the project being considered and even then encryption of identifiers is a preferable option. An example of this is in data linking, but even in such instances, best practice is to separate the creation of anonymous linkage variables as a process isolated from analysis or further sharing.

Treating the data

It is beyond the scope of this Guide to detail the various methods for treating data (that is, changing the data to decrease risks of disclosure), whether in aggregate or individual record form. There is much academic literature available as well as expertise in several Australian Government agencies (for example, Australian Bureau of Statistics, Australian Institute of Health and Welfare, Office of the Australian Information Commissioner, the CSIRO) and non-government entities (for example, universities) which can assist with specific advice on treating data to maintain privacy and confidentiality (for example, the De-identification Decision-Making Framework developed by the Office of the Australian Information Commissioner and CSIRO's Data61).

In brief, treatment methods include ‘data reduction’ treatments (which decrease the detail or variables available for sharing) or ‘data modification’ treatments (which change the values of individual data). However, both of these approaches may impact on the utility of the data. Whatever treatment is applied to the data, users will need to be made aware of what was done. It may not be desirable to detail exactly what treatment has been applied (as it may be that this information could be used to reverse the treatment), but general processes should be communicated to the users. This ensures that users understand the limitations of the data.

Another consideration is what other data may be available to a user. If a user is able to access other data in the same environment as the shared data, there may be a risk of the user comparing related or similar datasets (for example, two health datasets) in order to obtain more information than is available in the shared data. Ideally, this risk is controlled with the Project, People and Settings Principles, but if that is not possible, restricting the level of detail in the data shared, in order to protect against this risk, may be necessary.

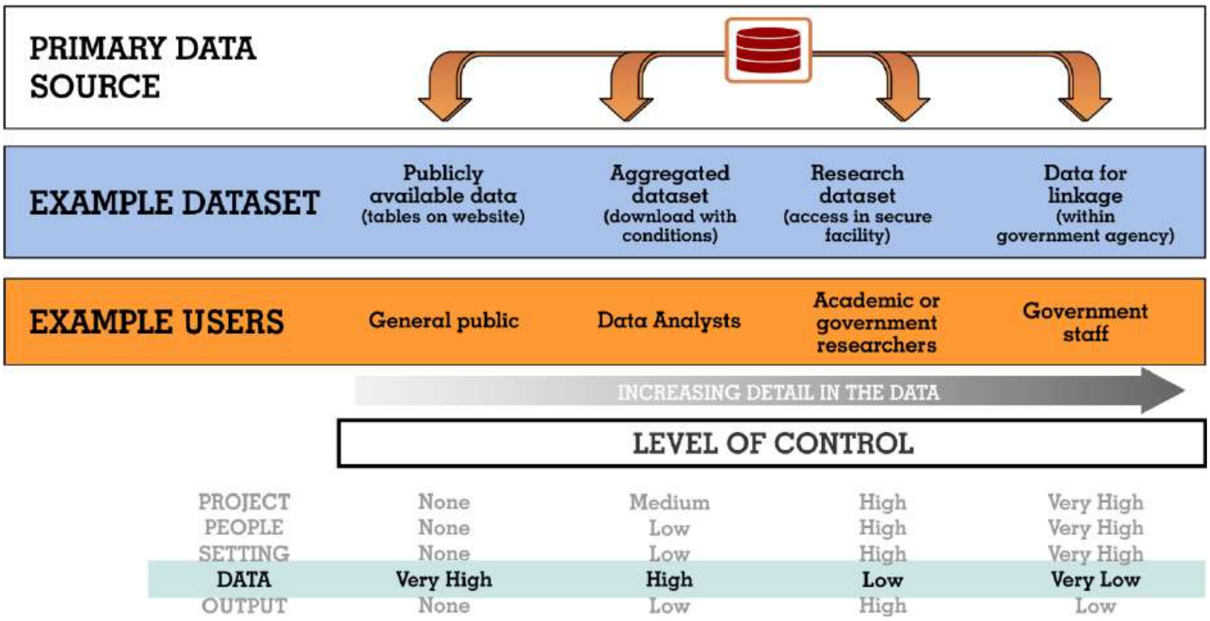
Whatever treatment is applied to the data, the aim is to balance disclosure risks with benefits (i.e. maximising the usefulness of the data). All data sharing carries some risk, however small, so it is better to focus on the likelihood of risks occurring and what can reasonably be done to reduce them. When applying the Data Principle, keep in mind that reducing detail in the data may reduce disclosure risks, but also reduce utility. If there is no reduction in benefit, then the reduced risk is always desirable.

Questions to ask: Data Principle

1. What risks can’t be controlled using the Project, People and Settings Principles?
2. Do direct identifiers need to be retained, for example, as a critical part of a project?
3. What further data treatment will be appropriate?
4. Are there specific issues associated with the sensitivity of the data (e.g. data which might identify where endangered species are located)?
5. How will data treatment affect utility of the data? How will this be communicated to authorised users?
6. What related data is expected to be available to the user in the same environment as the shared data? Can this be controlled?

Applying the Data Principle: Once Projects, People and Settings have been considered, it will be much clearer what controls are necessary around the Data Principle. The following diagram shows that a very high degree of treatment is necessary for data that is to be made public, while progressively less treatment is necessary as the detail in the data increases and access to the data is increasingly controlled by the other data sharing principles (for example, the level of authorisation required to access the data may be increased).

Diagram 5: the level of de-identification of data will depend on the level of detail being shared



PRIMARY DATA SOURCE

EXAMPLE DATASET

Publicly available data (tables on website)

Aggregated dataset (download with conditions)

Research dataset (access in secure facility)

Data for linkage (within government agency)

EXAMPLE USERS

General public

Data Analysts

Academic or government researchers

Government staff

	LEVEL OF CONTROL			
PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	None	Low	High	Very High
DATA	**Very High**	**High**	**Low**	**Very Low**
OUTPUT	None	Low	High	Low

5. Output Principle: The output from the data sharing arrangement is appropriately safeguarded before any further sharing or release

The Output Principle is concerned with what will happen to information or data created as a result of a data sharing arrangement. In many cases, this output will be a publication, report or other public release. Even if an output is not ostensibly made public (for example, a government program evaluation report), it is often advisable to assume that it will become so in the future (for example, under a Freedom of Information request) and treat it accordingly.

Sharing of data may result in the production of another dataset, which will then be shared further. For example, a data custodian provides a dataset to an expert data agency which improves or modifies the data and then provides authorised users with access to it for analysis. In this case, the expert data agency needs to conduct a new assessment using the Data Sharing Principles (in collaboration with the original data custodian) before the new dataset is on-shared.

When outputs are made public, or at least removed from an environment that had controls under the Settings Principle, a clear process for checking them is required. There are broadly two approaches to this – rules-based or principles-based.

Rules-based output checking

This is where simple deterministic rules (for example, minimum threshold of 100 observations contributing to a data point) are used to accept or reject outputs (either applied by the data custodian or by the user themselves). If data is supplied to a user for work in their own environment, it may be possible to provide a set of rules to guide appropriate outputs. Automated systems with simple rules mean users can usually obtain output results quickly, but this may well be at the cost of a loss of detail. These rules will tend to be conservative (focussing on preventing disclosure and not considering the utility of the output). As such, they can often block outputs that present no disclosure risk or may, depending on the exact thresholds, allow outputs to be released that should be withheld. Output checking systems relying on complex algorithms to assess for privacy risks can address many of these issues, but even these systems tend to be biased to eliminating all disclosure risks and therefore reduce the usefulness of the outputs.

Principles-based output checking

When outputs are requested from a highly controlled environment using detailed datasets, a principles-based approach is often more effective than using rules. For example, a simple principle may be that any output produced from a detailed dataset by a user does not present information about any particular unit represented in the dataset.

Principles-based output checking is an operational technique, usually seen as best practice, that ensures data outputs that are made public have a very low disclosure risk. Rather than preventing particular types or categories of outputs from being released, principles-based checking uses contextual information around the output as the basis for each assessment. Some thresholds can be put in place as examples of what would be expected to be acceptable content for an output. Because flexibility is built into the approach, these thresholds can be set quite strictly. If an output is likely to cross the threshold, the data custodian and users can discuss the proposed output so that a mutually agreeable compromise can be reached to maintain the safety of data sharing.

Principles-based output checking takes longer than a rule-based approach, but is better able to balance disclosure risk with usefulness of the output.

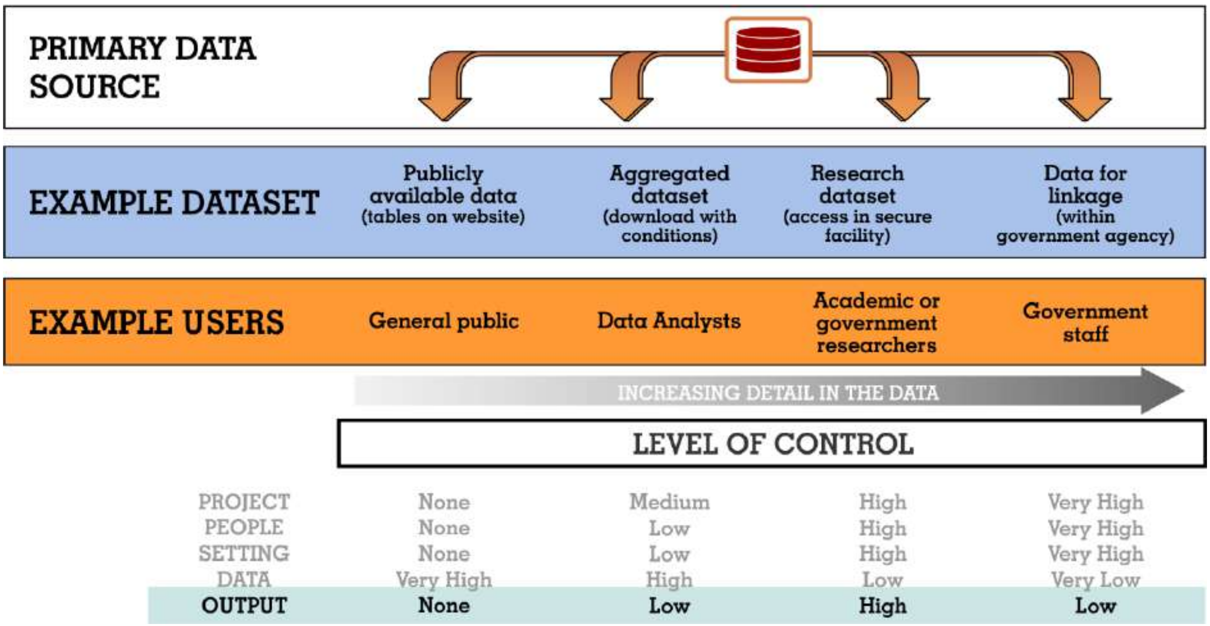
In order to function effectively, both users and data custodians need a clear and agreed understanding of who will conduct output checking and how a data output is to be checked. Some outputs may require special treatment to manage sensitivities. This often requires training of both groups and can be used to encourage users to produce outputs that require minimal effort to check.

Questions to ask: Output Principle

1. Will the output be released publicly or made available under new data sharing agreements?
2. Will simple rules meet the need of data custodians and users? How important is speed of response for the user?
3. Does the output adequately protect privacy and confidentiality?
4. Does output need to be checked and approved before being exposed to a broader audience?
5. What process will be followed for checking? Is it clear to the users?
6. Who will do the checking? Can the users check their own output? What resources are needed?
7. What extra checks need to be in place to account for sensitivities in the data?
8. Will people checking outputs or using the data need training on processes?

Applying the Output Principle: This principle is concerned with what happens to the results of the data use. The following diagram demonstrates that it is not possible to apply output controls to publicly available data, because the data custodian cannot control what that data will be used for. Use of more detailed data will require varying degrees of output controls. A research dataset may have an increased risk of an individual or organisation being identified in an output, and therefore increased protections will need to be applied to the output. Where a user is undertaking a linking project, very few protections may need to be applied given the intention may be to use the output in a new sharing arrangement (with the re-application of the Data Sharing Principles).

Diagram 6: the degree to which outputs will need protection will depend on the level of detail being shared



PRIMARY DATA SOURCE

EXAMPLE DATASET

Publicly available data (tables on website)

Aggregated dataset (download with conditions)

Research dataset (access in secure facility)

Data for linkage (within government agency)

EXAMPLE USERS

General public

Data Analysts

Academic or government researchers

Government staff

INCREASING DETAIL IN THE DATA

LEVEL OF CONTROL

PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	None	Low	High	Very High
DATA	Very High	High	Low	Very Low
OUTPUT	**None**	**Low**	**High**	**Low**

Part 3: After applying the Data Sharing Principles

Once the Principles have been applied, data custodians need to consider whether the controls appropriately safeguard the data to be shared. Data custodians need to ask: "Have the Principles reduced the risks of sharing to an acceptable level?" and "Can the data now be shared safely?" If the answer is no, then the data custodian can re-visit each of the Principles to adjust the level of control applied. If the risks of sharing cannot be reduced to an acceptable level, then that data should not be shared.

This discussion is more effective if the answer is supported by evidence rather than introducing theoretical risks that cannot be reasonably controlled. There may be more than one way to facilitate the data sharing arrangement - the controls applied under each of the Principles can be adjusted and re-assessed to enable safe sharing.

The data custodian also needs to consider whether the applied controls are proportionate to the data sharing arrangement. Applying the user-centric principle, it is important not to over engineer arrangements. If it is determined that the controls applied are excessive, consider reducing some controls to strike an appropriate balance between sufficient access, and the privacy or sensitivity of the data to be shared. An over-controlled solution will make it harder for users to do their work, with little corresponding improvement in the safety of the data.

The context for a data sharing arrangement may change over time and the application of the Principles may need to be reviewed for effectiveness through the life of a project. For example, a linkage project may mean that disclosure risks associated with the new linked dataset may be greater than the risks associated with each of the individual datasets that were linked, so risks may need to be controlled in different ways for each separate dataset. In addition, projects, users, organisations, technology and public expectations are all likely to experience change over time. What has worked in the past may not be appropriate under a future Government policy; and what is out of reach of current technologies may one day be commonplace and affordable. A clear review process should be built into all governance, reporting and assurance arrangements.

It is also important for data custodians to have clear processes for ensuring that all requirements of a Data Sharing Agreement are adhered to by authorised users. For example, this may include processes that assure the data custodian that any breaches of terms and conditions (such as unauthorised accesses or privacy breaches) have been reported and appropriate actions taken, or agreed destruction processes have been effectively carried out by authorised users at agreed times.

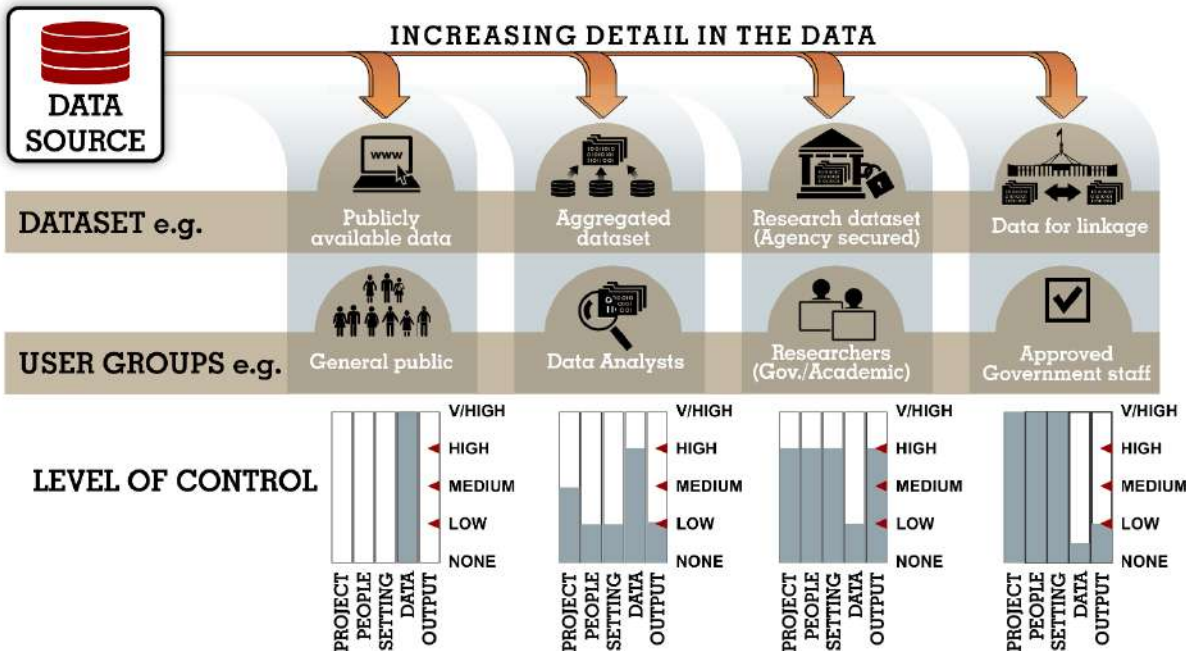
Efficient processes

Many of the controls that are put in place when using the Principles may be implemented as part of standard business processes. In order to save time for future requests, it is good practice to set up policies, processes, systems and teams which can be called on when required. For example, if a data custodian planned to share data with many academic users who will be engaged in the same kinds of research, the data custodian can pre-define and certify many of the processes and procedures. Then the only question to ask when a project comes in is "is this a legal and worthwhile project?" This will save time and money (although it is more work initially) and will mean the data custodian can be much more responsive to project requests.

Questions to ask: After applying the Principles

1. Have the controls minimised the risks associated with sharing the data?
2. If there are still unacceptable risks, which Principles need to be revisited?
3. What ongoing processes or governance are needed to monitor projects?
4. How are controls under each principle monitored to ensure they remain appropriate as circumstances change?
5. How can the application of the Principles be done more efficiently next time? Are the processes flexible enough to change if required?
6. What project reporting is needed, and to what person or body?
7. Is the data sharing process transparent and able to be audited?
8. Have end-of-life agreements specified in the project, such as a secure destruction process been adhered to?
9. Is the destruction process able to be observed or audited? Does it have a specified time by which it must happen?
10. Is there a well understood process for managing consequences if risks are realised (e.g. a confidentiality breach or unauthorised access)?

Diagram 7: Different controls applied to different datasets based on a primary data source



Further guidance

Expert advice on the application of these Principles and similar risk management frameworks is available from a number of Australian Government agencies such as the Australian Institute of Health and Welfare, the Australian Bureau of Statistics and the Office of the Australian Information Commissioner. This Guide operates in line with existing Government policy, legislation and technical advice.

Further useful information and guidance can be found at the following:

- The *Privacy Act 1988;*
- Australian Privacy Principles Guidelines;
- Guide to securing personal information;
- The *Archives Act 1983*;
- Commonwealth Risk Management Policy;
- De-identification Decision-Making Framework (OAIC and Data61);
- Confidentiality Series (ABS);
- High Level Principles for Data Integration;
- Australian Government Public Data Policy;
- Australian Government Information Security Manual;
- Productivity Report on Data Availability and Use;
- Australian Government's response to the Productivity Commission Data Availability and Use Inquiry;
- Secure Cloud Strategy;
- Australian Cyber Security Centre;
- NAA Information Governance; and
- NAA Building Interoperability advice (coming soon).

Where multiple jurisdictions contribute data, additional legislation and policies may apply.

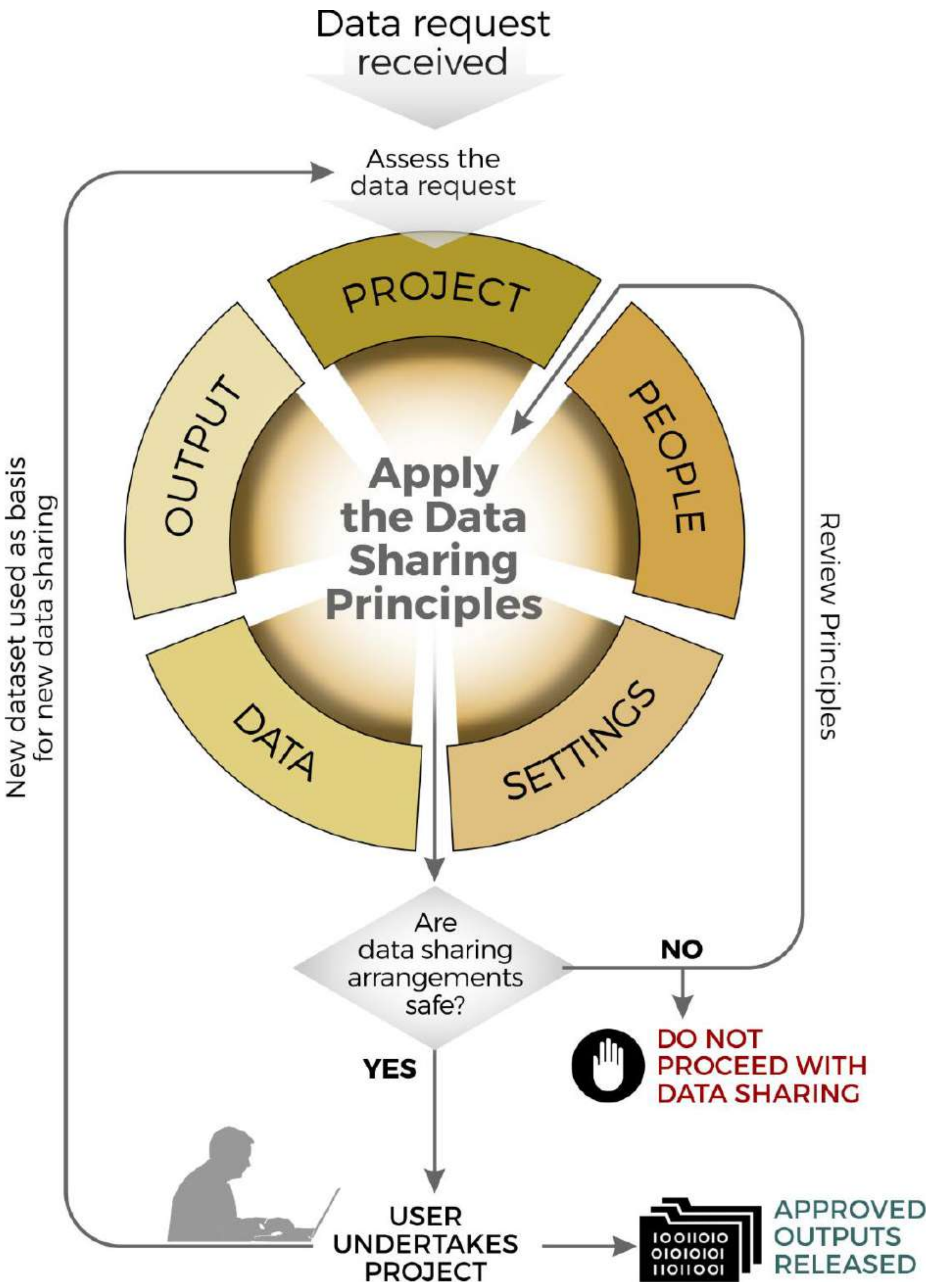
Appendices

Appendix A: Questions to ask under each Principle

Principle	Questions
Project	▪ Is the project in the public interest and does it satisfy a purpose test? ▪ Has all relevant information been provided to support assessment of the project proposal (e.g. who will access the data, for what purpose, over what period of time and what will happen to the data when the project ends)? ▪ What processes or governance arrangements are needed to assess, monitor and oversee the project? ▪ Who will make the assessment of whether to proceed with the project and do they possess the right capabilities to make the assessment? ▪ Are there any restrictions (e.g. legal or data custodian imposed restrictions) on how the shared data may be used? ▪ How will communication with applicants before and during the assessment of the project proposal be managed to maximise the likelihood of approval? What feedback will be provided? ▪ Does there need to be ethics approval from a governance body that considers the ethics of the proposal? ▪ Is consent from the original data providers required? ▪ What collaboration opportunities could the project provide to improve organisational processes?
People	▪ What, if any, process for authorisation is required for people or organisations to access data? Who facilitates this process? How long will any authorisation be valid? ▪ Will a legally binding undertaking or agreement to govern the access and use be required? Who needs to complete this undertaking or agreement? ▪ Does the user need to meet any specific criteria to access data (e.g. hold a current security clearance)? What are these criteria? ▪ Does the user have a history of good data handling practices? Does the user need to seek endorsement of their data handling practices from a responsible officer of their organisation? ▪ Does the user need to be trained in safe use, data storage and technical skills? Who develops and/or provides training? ▪ What sanctions (legal and non-legal) need to be available for misuse of data? Are these clear to the user? ▪ Are there any restrictions on who may apply to access the data (e.g. must be an Australian citizen, current affiliation with a particular research institution)?
Settings	▪ From what physical location(s) will the data be accessed? ▪ Does there need to be auditing/checks of these locations? ▪ What physical supervision is appropriate? ▪ What IT security needs to be in place? Will the security classification of the data influence IT security requirements? ▪ What electronic supervision as well as auditing/recording of use is available? ▪ Is certification of physical and/or IT environment necessary? If so, by whom? ▪ Do the controls limit misuse (by mistake and deliberate), interference, unauthorised access, modification, loss or disclosure? ▪ Do users understand how to access the data safely in the IT and/or physical environment? Is training required? ▪ How will data transfer into and out of a secure environment be managed?

Principle	Questions
Data	▪ What risks can’t be controlled using the Project, People and Settings Principles? ▪ Do direct identifiers need to be retained, for example, as a critical part of a project? ▪ What further data treatment will be appropriate? ▪ Are there specific issues associated with the sensitivity of the data (e.g. data which might identify where endangered species are located)? ▪ How will data treatment affect utility of the data? How will this be communicated to authorised users? ▪ What related data is expected to be available to the user in the same environment as the shared data? Can this be controlled?
Output	▪ Will the output be released publicly or made available under a new data sharing agreements? ▪ Will simple rules meet the need of data custodians and users? How important is speed of response for the user? ▪ Does the output adequately protect privacy and confidentiality? ▪ Does output need to be checked and approved before being exposed to a broader audience? ▪ What process will be followed for checking? Is it clear to the users? ▪ Who will do the checking? Can the users check their own output? What resources are needed? ▪ What extra checks need to be in place to account for sensitivities in the data? ▪ Will people checking outputs or using the data need training on processes?

Appendix B: Applying the Data Sharing Principles


Data request received
Assess the data request
PROJECT
PEOPLE
SETTINGS
DATA
OUTPUT
Apply the Data Sharing Principles
Review Principles
New dataset used as basis for new data sharing
Are data sharing arrangements safe?
NO
DO NOT PROCEED WITH DATA SHARING
YES
USER UNDERTAKES PROJECT
1001101010
0101010101
1101100101
APPROVED OUTPUTS RELEASED