



Rights, responsibilities and roles of data custodians

Introduction

1. This paper identifies the rights, responsibilities and roles of data custodians relative to those of the other key participants in data integration projects involving Commonwealth data for statistical and research purposes, namely integrating authorities and users of integrated datasets.

Who is a data custodian?

2. Data custodians are agencies responsible for managing the use, disclosure and protection of source data used in a statistical data integration project. Data custodians collect and hold information on behalf of a data provider (defined as an individual, household, business or other organisation which supplies data either for statistical or administrative purposes). The role of data custodians may also extend to producing source data, in addition to their role as a holder of datasets.

3. For any given data integration project (or family of projects)¹ involving Commonwealth data, there may be one or more data custodians. These may be from the same organisation or from separate institutions, will include at least one Commonwealth agency, and may include state/territory agencies and non-government organisations such as universities and private sector businesses.

Commonwealth data integration arrangements

4. Figure 1 is a stylised representation of the interactions between data custodians, integrating authorities, and data users around data integration arrangements that involve the use of at least one Commonwealth dataset for statistical and research purposes. The figure shows the key factors which data custodians need to consider before releasing data to an integrating authority, namely:

- Existing legislation which enables the release of, and access to, data by integrating authorities (i.e., legislation which applies to data custodians and integrating authorities).
- Privacy impacts relating to the use and disclosure of personal information or business data. This may include an assessment against the Commonwealth

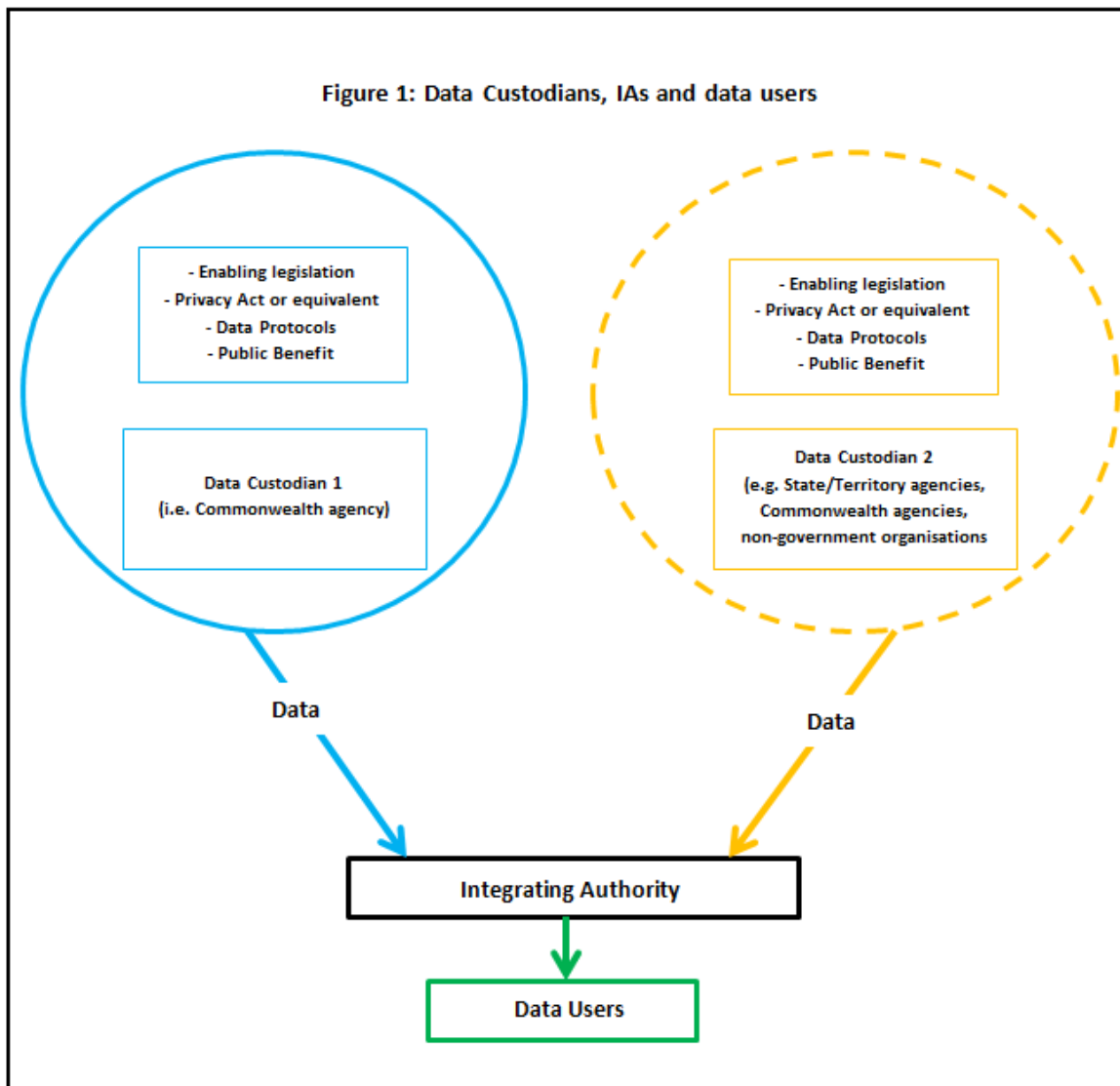
¹ A family of projects is defined as data integration projects using the same source datasets, for similar purposes, with the same integrating authority and these are treated as a single program for the purposes of the approval process. References to data integration projects in the remainder of this document include families of projects.

Privacy Act and equivalent state/territory or other legislation. Agency-specific legislation will also need to be considered.

- The existence of any data protocols governing access to, and the use of, datasets. An example is the need for ethics committee approval for projects using human research data (e.g., clinical trials, population health research and health services data).
- Ensuring that the data integration project takes into account the public benefit which can be derived from the use of integrated datasets.

5. The choice of integrating authority will be based on a consultative process led by the data custodians taking into account any preferences of the data users. Data custodian need to give *in principle* approval for the project to proceed before an integrating authority is appointed. Data custodian(s) remain individually accountable for the source data used in statistical data integration projects (refer to the Commonwealth's Statistical Integration Principle 2 – Custodian's Accountability) and must ultimately be satisfied with the chosen integrating authority.

6. Commonwealth data custodians are required to address all of these factors. The requirements will vary among different data custodians as indicated in Figure 1.



The rights and responsibilities of data custodians

7. Data custodians have a number of ‘rights and responsibilities’ related to the release of, and access to, source datasets for data integration projects. These form the basis for how data custodians will work collaboratively with other participants involved in data integration projects.

8. The key rights and responsibilities of data custodians in relation to integrating authorities are listed below.

- It is the responsibility of data custodians to ensure they are authorised by legislation or consent to release identifiable data to an integrating authority. The data custodians must also be satisfied that the integrating authority has the necessary legislative protections in place prohibiting disclosure of identifiable data, before the commencement of a data integration project.

- If a data custodian approves a project, the data custodian will need to provide data to the relevant integrating authority for that project. Data will be protected and safely managed by the integrating authority throughout the project life cycle and in accordance with any requirements of data custodians.
- It is the responsibility of data custodians to ensure good data management practices, (including clear documentation, the use of standard definitions and classifications, the maintenance of appropriate metadata, and quality assurance) are undertaken before data is provided to an integrating authority.
- It is the right of data custodians to have data linkage, merging and access services provided on their behalf by an integrating authority.
- Where the Cross Portfolio Data Integration Oversight Board advises on amendments to ‘high risk’² projects (or where a concern is raised), data custodians, the integrating authority and data users will need to collaborate on how to make improvements to such project(s).
- Data custodians may collaborate with integrating authorities on the content of training material provided to data users. Input, advice and assistance will be provided at the discretion of data custodians.
- Data custodians are responsible for consulting with the integrating authority about the information the integrating authority will provide when registering the project on the Public Register of Data Integration Projects.

9. Data custodians also have rights and responsibilities in relation to data users.

- Data custodians are responsible, along with the integrating authority for consulting with data users on any material changes or updates to a data integration project (regardless of whether changes originate from data custodians or integrating authorities). This will occur before data users start examining integrated datasets. Possible issues raised by integrating authorities may include the technical feasibility of the project or the limitations of data use.
- Data custodians may provide input, advice and assistance to integrating authorities on the content of training material.
- Data custodians can expect that data users are aware of, and understand, sanctions which apply for attempts to identify (or re-identify) individuals or organisations; disseminating outputs that enable the identification of individuals or organisations; or the misuse of data.
- It is the right of data custodians to approve or disapprove project proposals, in whole or in part (this decision could be influenced by the technical feasibility assessment made by the integrating authorities). Data custodians also have the right to prioritise and schedule data extraction work associated with project proposals, taking account of the range of data extraction requests that may be

² Data custodians assess the risk of a project in accordance with the [risk assessment framework](#).

outstanding at the time. Data custodians are responsible for informing data users of these outcomes.

- Data custodians may collaborate with data users and integrating authorities on how to make improvements to ‘high risk’ project(s), based on advice provided by the Cross Portfolio Data Integration Oversight Board.
- Data custodians have a right to implement cost recovery policies (to recover the costs of preparing and extracting datasets, for example) for data integration projects. *Note: cost recovery by Commonwealth agencies must comply with the Australian Government Cost Recovery Guidelines (July 2005) published by the Department of Finance and Deregulation.*
- Data custodians and data users are responsible for ensuring that datasets are used for the approved purposes only. This is facilitated by practices which help avoid the misinterpretation of data. Examples include the supply of appropriate metadata by data custodians and the testing of assumptions made in respect of the data by researchers.

The role of data custodians in data integration projects

10. Data custodians have six key roles in the Commonwealth data integration arrangements. These roles reflect the need for data custodians to strike a balance between maximising the inherent value of data assets and minimising privacy concerns associated with the use of this data. The roles are:

- Safe storage of unit record level information;
- Assessing the level of risk for each data integration project;
- Ensuring compliance with relevant legislation, including privacy, for data release;
- Entering into agreements with integrating authorities;
- Safe transmission of data; and
- Maximising the value of data holdings.

(1) Safe storage of unit record level information

11. Data custodians must have policies and procedures in place which contain information on how data custodians will interact with integrating authorities and data users, along with the rights and obligations that exist, to ensure the safe storage of unit record level information. Examples include communication, information security, training and governance strategies.

12. The safe storage of unit record data should be considered along with quality assurance. Data custodians must ensure, as far as practicable, the accuracy, currency and timeliness of data supplied to an integrating authority. The need to ensure data quality aligns with the high level statistical integration principles, which specify the importance of data custodians following good data management practices and maintaining the quality attributes of data.³ For personal information, it is also consistent with Information Privacy Principle 8 of the *Privacy Act 1988* which states the record-keeper (i.e. the data custodian) is responsible for checking the accuracy and completeness of personal information before it is used. Equivalent principles for business data should also be considered.

13. Quality assurance is an essential step to help minimise any potential problems which may arise with integrated datasets. Good data management practices include the provision of clear documentation, the use of standard definitions and classifications, and the maintenance of appropriate metadata, including quality attributes of the data. Any quality (or software) issues arising from the use of integrated datasets will be addressed by a governance protocol developed by data custodians and integrating authorities (see Role 4: Entering into agreements with integrating authorities).

(2) Assessing the level of risk for each data integration project

14. A key role of data custodians is to determine the level of risk for a data integration project, using the risk assessment framework developed for Commonwealth data integration projects. The level of risk is an important part of determining if a project should proceed and whether an accredited Integrating Authority is required to manage the integration project (i.e. if the project is assessed as high risk). Where there is more than one data custodian, a lead data custodian may be appointed by the data custodians to compile the risk assessments.

(3) Ensuring compliance with relevant legislation, including privacy, for data release

16. Data custodians need to take into account the potential privacy impacts for any given data integration project. Guidelines for the collection, use and disclosure of personal information by Commonwealth and ACT government agencies are stipulated in

³ Commonwealth's Statistical Integration Principle 1 – Strategic Resource: Responsible agencies should treat data as a strategic resource and design and manage administrative data to support their wider statistical and research use.

Information Privacy Principles contained in the *Privacy Act 1988*. The same legislation also extends to some private sector organisations and small businesses (including non-profit organisations or unincorporated associations) under National Privacy Principles.⁴

17. Agency-specific legislation also affects the disclosure of personal information obtained through data collections. Examples include secrecy provisions that apply to statistical collections under the *Census and Statistics Act 1905* (ABS), identifiable data disclosed under the *Health Insurance Act 1973* and *National Health Act 1953* (Department of Health), and the disclosure of protected information relating to income support (*Social Security (Administration) Act 1999*) and family payments (*A New Tax System (Family Assistance) (Administration) Act 1999*) (Department of Social Services).

18. For all data integration projects, data custodians must determine whether they are authorised to release identifiable data to an integrating authority either by the data custodian's legislation or by consent from the data provider (that is, the person, family, household, business or other organisation who originally supplied the data for statistical and administrative purposes).

19. The data custodian must be satisfied that the integrating authority has the necessary legislative protections in place prohibiting disclosure of identifiable data, other than where allowed by law. In particular accredited Integrating Authorities undertaking high risk data integration projects must be bound by the Commonwealth Privacy Act (or a state/territory equivalent) and be subject to criminal penalties for a breach of legislation with regard to an unauthorised disclosure of information. For low and medium risk projects, at a minimum, integrating authorities must have an appropriate policy framework in place to ensure that no identifiable data is disclosed, other than where allowed by legislation⁵.

20. A Privacy Impact Assessment should be considered for 'high risk' projects. This will help data custodians identify and address any potential privacy risks around the collection, use and release of data. An example of work undertaken as part of a Privacy Impact Assessment is an examination of the public interest test by health data custodians. Under Section 95a of the *Privacy Act 1988*, National Health and Medical Research Council (NHMRC) guidelines allow for the use and disclosure of health information where it substantially outweighs the public interest in maintaining privacy.

21. Where legislative authority does not exist to release data to an integrating authority, informed consent must be obtained from data providers before the release of data to an integrating authority.

⁴ The *Privacy Act 1988* covers small businesses and private sector organisations with a turnover of greater than \$3 million. Businesses with turnover of less than \$3 million are not covered, although some exceptions may apply.

⁵ For 'high risk' projects, data custodians must be assured that the nominated integrating authority has the necessary legal protections in place. For medium risk projects, integrating authorities must have an appropriate policy framework in place to ensure that no identifiable data is disclosed, other than where allowed by legislation.

(4) Entering into agreements with integrating authorities

22. Each data custodian must enter into an agreement with a nominated integrating authority. This agreement may take the form of a contract, Memorandum of Understanding or other arrangement as appropriate for the parties concerned. When the data custodian and integrating authority is the same agency, appropriate internal governance arrangements, rather than an agreement, will need to be in place. The purpose of a project agreement is to help ensure that datasets are managed and used in accordance with data custodian requirements throughout the life of the project.⁶

23. The terms of the agreement will vary on a project-by-project basis, but will generally consist of core elements such as:

- Information on how data will be safely managed, including the provision of secure data arrangements by integrating authorities, to help manage project risks;
- The use of data protocols that balance risk and public benefit (e.g., the use of ethics committees for human-based health research);
- Specifying control mechanisms, in collaboration with integrating authorities, to assess and ensure that individual or business data is not likely to be identified. This may take the form of data modification or data reduction techniques.
- Developing governance protocols to investigate and resolve software issues, along with any anomalies, outliers and data quality issues not previously identified or which arise from the creation of new integrated datasets. Given that data custodians have a key role in quality assurance, it is expected that the data supplied to integrating authorities will be of high quality. The protocol will specify how data custodians will work with integrating authorities and data users. Data custodians should always consider intellectual property rights when deciding whether they are able to make data available for a particular project that would involve using any externally owned software or other technology for transmission of data to the integrating authority or allowing the integrating authority to use such software.
- Specifying any special conditions which must be adhered to by data users. This may include, for example, training requirements and the signing of confidentiality agreements with data custodians. It may also include assurances that data users will make valid use of datasets. For example, data users may be required to seek clearance from data custodians on the use and interpretation of data before publishing research outputs. However, this is not a uniform requirement across Commonwealth agencies.
- The use of communication, technology, training and other processes to minimise the risk of identification of individuals or businesses.

⁶ Commonwealth's Statistical Integration Principle 2 – Custodian's Accountability: Agencies responsible for source data used in statistical data integration remain individually accountable for their security and confidentiality.

(5) Safe transmission of data

24. A key function of data custodians is to ensure the safe transmission of data to integrating authorities. The safe transmission of data should be undertaken in accordance with legislative and policy requirements prior to the commencement of data linkage operations, and in accordance with project agreements. The *Australian Protective Security Policy Framework* provides further information on the safe transmission of Commonwealth data.

(6) Maximising the value of data holdings

25. Data custodians must seek to maximise the value of any administrative datasets that are collected and take into account the public benefit which can be derived from statistical and research proposals submitted by data users, as per Commonwealth Statistical Principle 1 (treat data as a strategic resource). However, Commonwealth administrative data cannot be used for statistical or research purposes if this contravenes legislation or any commitment made to data providers regarding the purpose for which their data may be used, or the data is commercial in confidence.

26. Cross-Commonwealth data integration projects are particularly useful for informing whole-of-government policy perspectives and helping to make the best possible use of data that already exists and where possible, minimising respondent burden around data collections. Two current examples of such data integration projects are listed below.

- The Longitudinal Study of Australian Children (LSAC) links one study period to the next. It also links to administrative databases and aggregate census data as a supplement to information collected in the survey. LSAC aims to help improve the understanding of factors influencing childhood development and collects information on children's physical, cognitive and emotional development. This will help guide policies and interventions to address issues affecting the development and wellbeing of Australian children.
- The Business Longitudinal Database combines small and medium size business characteristics (sourced from the ABS) with financial data from the Australian Taxation Office and the Australian Customs and Border Protection Service. The project provides a basis for measuring the performance over time, as well as key business drivers.

Multiple roles of data custodians in data integration projects

27. For some data integration projects, it is possible that data custodians may have multiple roles where a data custodian may also be the data user (e.g., a Commonwealth agency) and/or the integrating authority.

28. When an entity has more than one role, appropriate internal governance and project documentation, consistent with the Commonwealth principles and governance arrangements for data integration should be in place.

More information

29. Any questions about the roles and responsibilities of data custodians should be emailed to statistical.data.integration@nss.gov.au