



Australian Government

# High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes

Endorsed by  
Portfolio Secretaries  
**3 February 2010**

An initiative of the  
**Cross Portfolio Statistical Integration Committee (CPSIC)**

## About CPSIC

Australian Government Portfolio Secretaries have established a Cross Portfolio Statistical Integration Committee (CPSIC), jointly chaired by DoHA and ABS, to create an Australian government approach to facilitate linkage of social, economic and environmental data for statistical and research purposes.

Cross Portfolio Statistical Integration Committee members are: Attorney-General's Department, Australian Bureau of Agricultural Resource Economics, Australian Bureau of Statistics, Australian Public Service Commission, Department of Broadband, Communications and the Digital Economy, Department of Defence, Department of Health and Ageing, Department of Education, Employment and Workplace Relations, Department of the Environment, Water, Heritage and the Arts, Department of Finance and Deregulation, Department of Foreign Affairs and Trade, Department of Human Services, Department of Immigration and Citizenship, Department of Innovation, Industry, Science and Research, Department of Infrastructure, Regional Development and Local Government, Department of Families, Housing, Community Services and Indigenous Affairs, Department of the Prime Minister and Cabinet, Department of Resources, Energy and Tourism, Department of the Treasury and the Department of Veteran's Affairs.

## Statistical Integration – why?

Statistical data integration involves integrating unit record data from different administrative and/or survey sources to provide new datasets for statistical and research purposes. The approach leverages more information from the combination of individual datasets than is available from the individual datasets taken separately. Statistical integration aims to maximise the potential statistical value of existing and new datasets, to improve community health, as well as social and economic wellbeing by integrating data across multiple sources and by working with governments, the community and researchers to build a safe and effective environment for statistical data integration activities.

Integrated datasets provide public benefits in terms of improved research, supporting good government policy making, program management and service delivery. Integrated datasets also create an important opportunity to expand the range of official statistics to better inform Australian society.



# 1

## PRINCIPLE ONE – 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.**

**This principle aims to maximise statistical and research use of existing and new Commonwealth data sets.**

Administrative data represent a public asset that requires protection and management for appropriate purposes. When designing and managing administrative datasets, the responsible agency should consider the potential statistical value of the datasets for public good, both in terms of use by their own agency, and use more broadly.

Administrative data cannot be used for statistical purposes if this contravenes legislation or any commitment made to data providers, or the data is commercial in confidence. Nor should it be used for statistical purposes if this use clearly threatens the integrity of the administrative data.

Where administrative data is likely to have high public value for statistical use, those providing data should be informed of the potential for statistical use at the time of data collection. Where historical data has been collected without providing this information it should still be considered for statistical use, but not if this is prohibited by commitments made to providers at the time of collection.

Where administrative data is likely to have value for statistical use, efforts should be made to maximise that value through good data management, including the use of standard definitions and classifications and the maintenance of appropriate metadata, including quality attributes of the data.

Where data is sought for statistical purposes, consideration should be given to using existing administrative sources in preference to imposing additional load on providers through the institution of a new statistical collection.

The statistical and research value of administrative data should be maximised, within legal and practical constraints, by granting broad access for research purposes to data that is not likely to enable identification. Commonwealth administrative data should not generally be withheld from research for reasons of Intellectual Property.

## PRINCIPLE TWO – Custodian’s Accountability

**Agencies responsible for source data used in statistical data integration remain individually accountable for their security and confidentiality.**

**This principle ensures that data custodians recognise their continued accountability for their data within integrated datasets and establish adequate controls over the use of personal or other sensitive data in data integration projects.**

Each responsible agency for source data:

- must agree mechanisms to achieve adequate control and manage risk appropriate to their own situation. These mechanisms may include the use of particular institutional arrangements with trusted institutions, the use of specified standards and audits against those standards, and the potential application of sanctions.
- will need to agree the nature of valid uses that can be made of the integrated datasets and the approval mechanism to be applied to applications to use the datasets, as well as any control mechanisms to be applied to such use.
- will need to manage the potential increase in identifiability of data for which they are responsible when it is used in conjunction with data from other sources. It will need to agree mechanisms by which it can assure itself that outputs from the statistical data integration are not likely to enable the identification of individuals or businesses.
- will need to agree the final content of any new data integration proposal, or any material change to an existing data integration proposal as part of the approval process. They must be kept informed of, and agree, more minor proposed changes to existing proposals.

Where an agency does not agree to the use of its source data in a statistical integration proposal, that data will not be included. For example this might occur if the proposal threatens the integrity of the administrative data.

# 3

## PRINCIPLE THREE – Integrator’s Accountability

**A responsible ‘integrating authority’ will be nominated for each statistical data integration proposal.**

**This principle sets out the responsibilities of integrating authorities to manage the data integration project from start to finish in line with the agreements made with data custodians and requirements as part of approval processes.**

An integrating authority must be identified for each statistical data integration proposal. This authority will be held responsible for the sound conduct of the statistical data integration proposed, in line with the agreed requirements of the responsible agencies.

Although the integrating authority is the single organisation ultimately accountable for the Statistical Data Integration project, it may work with a network of agencies to achieve the data integration. For example, it might use another agency to undertake linkage or to support dissemination.

The integrating authority will ensure appropriate governance is in place including:

- an open approval process is followed;
- documentation of the proposal;
- the impact on privacy;
- risks have been assessed, managed and mitigated
- the expected costs and benefits; and
- the outputs.

A family of data integration projects using the same source datasets, for similar purposes, with the same integrating authority, may be treated as a single program for the purposes of the approval process.

The integrating authority will be responsible for the ongoing management of the integrated data, ensuring it is kept secure, confidential and fit for the purposes for which it was approved.

If it is an ongoing project, the integrating authority will be responsible for initiating and managing its regular review, in consultation with source data agencies.



## 4

### PRINCIPLE FOUR – Public Benefit

**Statistical integration should only occur where it provides significant overall benefit to the public.**

**This principle ensures there is a demonstrated ability to produce significant outputs from the integrated dataset and an independent assessment is made that the public good outweighs the privacy imposition and risks to confidentiality.**

There should be a demonstrated ability to produce significant outputs from the integrated dataset.

There should be an independent assessment of the balance of public good against the privacy imposition and risks to confidentiality. Examples include community representation on the steering committee, the use of an ethics committee, or the use of an advisory committee with community representation and the ability to report independently of the agencies involved in the proposal.

Ongoing programs should be reviewed on a three yearly basis to ensure a continuing overall benefit.

## 5

### PRINCIPLE FIVE – Statistical & Research Purposes

**Statistical data integration must be used for statistical and research purposes only.**

**This principle requires that where data integration is approved and implemented for statistical and research purposes, it is not then used for regulatory purposes, compliance monitoring, or service delivery. This helps to ensure that the risk of breaches of personal information and the potential impact of any inadvertent breach remain low.**

Statistical data integration must not be used for non-statistical purposes requiring the identification of an individual person, household, family or business, for example, the delivery of services to particular individuals, individual compliance monitoring, client management, incident investigation, or for regulatory purposes. However the insights gained through statistical and research outputs are expected to improve processes in these areas.

There must be no feedback of information relating to individuals or individual businesses from the statistical data integration project back to the originating administrative sources, unless that feedback was derived from a single source and is returning the same data to that source.

## PRINCIPLE SIX – Preserving Privacy & Confidentiality

**Policies and procedures used in data integration must minimise any potential impact on privacy and confidentiality**

**This principle ensures privacy and confidentiality are preserved to the maximum extent possible.**

Operational, administrative and personal identifiers should be removed from datasets as soon as they are no longer required to meet the approved purposes of the statistical data integration project.

Where identifiers need to be retained, for example for longitudinal studies, they should be kept separate from the integrated dataset.

The number of unit records and data variables to be included in an integrated dataset should be no more than required to support the approved purposes.

The type of matching used (exact, probabilistic or statistical) should be chosen as the minimum needed to support the approved purposes, and the range of attributes used to establish a common identity should be the minimum necessary for the linking operation to succeed.

Access to potentially identifiable data for statistical and research purposes, outside secure and trusted institutional environments should only occur where: legislation allows; it is necessary to achieve the approved purposes; and meets agreements with source data agencies.

Risks of indirect as well as direct identification should be carefully managed when data is disseminated outside secure and trusted institutions, particularly in terms of units with unusual characteristics. This management must take account of the potential increase in identifiability of one set of data when combined with another set. It might involve strict data use licensing conditions, reducing detail, perturbing data, or seeking the consent of the individual or business involved to release potentially identifiable data, the last of these being most likely in the case of business data.

Once the approved purpose of the project is met, the related datasets should be destroyed, or if retained, the reasons for and necessity of retention documented, and a review process set up. If such retention was not part of the initial approval process, re-approval of the decision to retain is required.

Archiving of statistically integrated data sets should be restricted to confidentialised datasets.

## PRINCIPLE SEVEN – Transparency

**Statistical data integration will be conducted in an open and accountable way.**

**This principle ensures the public is aware of how Commonwealth government data is being used for statistical and research purposes.**

The main elements will be:

- governed in an open accountable way
- ensure stakeholders and the community are kept informed of any statistical data integration project undertaken, by publishing appropriate details of the project such as the datasets, the purpose, provision for access, use made of the dataset, the make up and role of any advisory body or steering group, the role of involved institutions, the approval process, and the review process.
- appropriate privacy impact assessment
- each project is subject to audit, agencies responsible for source data and data integration, will agree on audit schedules.

### Resources

“A Good Practice Guide to Sharing your Data with Others”

From the National Statistical Service (NSS) website:

<http://www.nss.gov.au/nss/home.NSF/pages/NSS+Resources>

For more information, please contact the CPSIC Secretariat by:

Phone: (02) 6252 7198

E-mail: [inquiries@nss.gov.au](mailto:inquiries@nss.gov.au)