



Australian Government

Office of the Australian Information Commissioner

Guidelines for Data-matching in Australian Government Administration

Draft as at December 2011

Contents

Guidelines for Data-matching in Australian Government Administration	1
Key terms	1
Background	3
The purpose of the Guidelines	3
Who should use the Guidelines?	3
Status of the Guidelines	3
History of the Guidelines.....	4
Role of the OAIC.....	5
Investigations	5
Data-matching.....	5
Application of the Guidelines.....	6
What scale of data-matching do the Guidelines apply to?.....	6
Reporting and notification requirements where the Guidelines do not apply	6
Guideline 1 – Application of the Guidelines	8
Similar data-matching programs to be treated as a single data-matching program.....	8
Data-matching programs involving more than one agency	9
Guideline 2 – Deciding to carry out or participate in a data-matching program	10
Guideline 3 – Prepare a program protocol	11
Purpose of the program protocol	11
What should the program protocol contain?	11
Program protocols for similar data-matching programs treated as a single program.....	12
Publishing the program protocol	12
Changes to data-matching programs.....	13
Guideline 4 – Prepare a technical standards report	14
Purpose of the technical standards report	14
What should the technical standards report contain?	14
Comply with the technical standards report	15
New data-matching programs	15
Changes to data-matching program specifications	15
Guideline 5 – Notify the public	16
Obligation to notify	16

Content of public notice.....	16
Forms of public notice.....	16
Submission for inclusion in the Personal Information Digest.....	17
Guideline 6 – Notify individuals of proposed administrative action	19
Guideline 7 – Destroy information that is no longer required	21
Guideline 8 – Do not create new registers, data-sets, or databases.....	22
Guideline 9 – Regularly evaluate data-matching programs	23
Guideline 10 – Seeking exemptions from Guideline requirements	25
Explaining the public interest grounds	25
OAIIC response to advice	26
Publication of advice	26
Guideline 11 – Data-matching with entities other than agencies.....	27
Guideline 12 – Enable review by the OAIIC	28
Appendix A: Content of data-matching program protocols.....	29
Description of the data-matching program	30
Reasons for deciding to conduct the data-matching program.....	31
(k) Alternative measures to data-matching that were considered, and the reasons why they were rejected. (Alternative methods	32
(l) Information about any pilot testing of the program. (Pilot programs	32
(m) A statement of the costs and benefits of the	33
Appendix B: Technical standards report.....	34
Purpose	34
Contents	34
Appendix C: Statement of costs and benefits for data-matching programs.....	37
Introduction	37
Purpose of estimating costs and benefits.....	37
Methods of presenting cost/benefit information.....	38
Methods of presenting information	38
Estimating costs	40
Estimating benefits	42

Key terms

Administrative action means action taken in response to a match obtained through a data-matching program that materially affects any individual or class of individuals, including, but not limited to:

- any action directly detrimental to an individual, such as reducing a benefit or imposing a penalty
- the initiation of an investigation which might lead to action directly detrimental to the individual the subject of the investigation, and
- the disclosure of information to a third party, where the disclosure might cause harm (including embarrassment) to the individual to whom the information relates.

Agency has the meaning set out in s 6 of the Privacy Act and includes, amongst other things, a Minister, an Australian Government Department, an ACT Government Department, and a Norfolk Island agency.

ANAO means the Australian National Audit Office.

Commissioner means the Information Commissioner within the meaning of the *Australian Information Commissioner Act 2010*.¹

Database means a structured collection of data or records, stored by means of a computer in a manner that facilitates retrieval.

Data-matching means the bringing together of data-sets that contain personal information, and that come from different sources, and the comparison of those data-sets with the intention of producing a match.

Data-matching cycle means a single iteration of all the steps and processes involved in a data-matching program.

Data-matching program means the conduct of data-matching, in accordance with clearly defined steps and criteria, to assist one or more agencies to achieve a specific objective. A data-matching program may involve more than one data-matching cycle.

Data-set means a discrete, ordered collection of data. A data-set may be sourced from a database, and may be defined by specific criteria, for example, the receipt of a certain benefit within a given period.

FOI Act means the *Freedom of Information Act 1982* (Cth).²

Guidelines means Guidelines 1–12 as set out in this document.

¹ See www.comlaw.gov.au/Details/C2010A00052

² See www.comlaw.gov.au/Series/C2004A02562.

IPPs means the Information Privacy Principles set out in s 14 of the Privacy Act which apply to agencies.

Match means a result produced by data-matching, including a meaningful discrepancy, in relation to which administrative action may be taken by the matching agency, or source agency or organisation.

Matching agency means, in relation to a data-matching program, the agency whose facilities or resources are used to conduct a data-matching program.

OAIC means the Office of the Australian Information Commissioner.

Organisation has the meaning set out in s 6C of the Privacy Act and, in general, includes all businesses and non-government organisations with an annual turnover of more than \$3 million, all health service providers and a limited range of small businesses (see ss 6D and 6E of the Privacy Act).

Personal information has the meaning as set out in s 6 of the Privacy Act:

... personal information means information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.

Primary user agency means, where a data-matching program involves more than one agency, the agency that makes the most use of the results of a data-matching program.

Usually the primary user agency will also be the matching agency, but there may be data-matching programs where the matching agency does not use, or uses only to a limited extent, the results of the data-matching program. Where there is more than one agency using the results of a data-matching program, those agencies should agree which is the primary user agency.

Privacy Act means the *Privacy Act 1988* (Cth).³

Source entity means any entity, including an agency or organisation, that discloses a data-set containing personal information to a matching agency for use in a data-matching program.

Source data means the record, including electronic or paper records, from which information (held in a database or data-set) has been provided for use in a data-matching program.

Technical standards report means a report of the kind described in Appendix B.

User agency means an agency that uses the results of a data-matching program.

³ See www.comlaw.gov.au/Details/C2011C00503.

Background

The purpose of the Guidelines

1. The Guidelines have been developed to assist Australian Government, Australian Capital Territory (ACT) Government, and Norfolk Island Government agencies to use data-matching as an administrative tool in a way that complies with the IPPs and the Privacy Act, and is consistent with good privacy practice.

Who should use the Guidelines?

2. This document should be used by agencies that handle personal information and wish to use data-matching to determine whether administrative action is warranted.
3. These Guidelines do not apply to data-matching where Tax File Numbers are used. The *Data-matching Program (Assistance and Tax) Act 1990* (Cth)⁴ regulates the use of Tax File Numbers in comparing personal information held by the Australian Taxation Office and by certain ‘assistance agencies’ including the Department of Human Services (which administers the Centrelink, Child Support Agency, and Medicare Programs) and the Department of Veteran’s Affairs.⁵ The OAIC has issued separate mandatory guidelines in respect of the data-matching programs authorised by that Act.⁶
4. Further, the Privacy Commissioner has issued mandatory guidelines⁷ under s 135AA of the *National Health Act 1953* (Cth)⁸ that regulate the storage, use, disclosure, and linkage of patient claims information collected under the Pharmaceutical Benefits Scheme and the Medicare program. Agencies dealing with Medicare and Pharmaceutical Benefits Scheme information should consider their obligations under those guidelines before matching that information.

Status of the Guidelines

5. This document, including the Guidelines, is provided by way of guidance, and represents the OAIC’s view on best practice with respect to data-matching.
6. As such, compliance with the Guidelines is voluntary.

⁴ See www.comlaw.gov.au/Details/C2006C00591.

⁵ Section 3 of the *Data-matching Program (Assistance and Tax) Act 1990* (Cth) provides that **assistance agency** means:

- (a) the Department of Health and Family Services; or
- (b) the Department of Employment, Education and Training; or
- (c) the Department of Social Security; or
- (d) the Department of Veterans’ Affairs; or
- (e) the Human Services Department.

⁶ See www.privacy.gov.au/materials/types/download/8687/6526.

⁷ See www.comlaw.gov.au/Details/F2008B00554.

⁸ See www.comlaw.gov.au/Details/C2011C00638.

7. The OAIC encourages agencies to agree to adopt this document and comply with the Guidelines. However, an agency that has so agreed would not be acting unlawfully if it did not comply, unless the acts or practices of the agency constitute a breach of the Privacy Act.
8. The IPPs regulate the way in which agencies handle personal information. Under the Privacy Act, the OAIC has the power to audit agencies, investigate complaints, or investigate on the Commissioner's initiative to determine whether agencies are complying with the IPPs (see below at 'Role of the OAIC').
9. The OAIC may take the Guidelines into account when assessing whether an agency has complied with the IPPs. However, this document provides guidance on how to comply with the IPPs when carrying out data-matching activities, and guidance on other additional matters that may not be covered by the IPPs. Accordingly, a breach of the Guidelines will not necessarily constitute a breach of the IPPs.
10. The OAIC considers that the adoption of this document supports good privacy practice, reflects a commitment to the protection of individual privacy and promotes an Australian society in which privacy is respected.
11. The OAIC publishes information about Government data-matching activities in its annual reports, including a summary of program protocols received by the OAIC in each financial year.

History of the Guidelines

12. In 1990, the Privacy Commissioner issued a consultation draft of the *The use of data matching in Commonwealth administration – Guidelines* (former guidelines) for public comment.
13. In 1992, the Privacy Commissioner released the former guidelines for adoption by agencies.
14. The former guidelines were subsequently revised and re-issued in November 1995, and again in February 1998.
15. This document implements changes required as a consequence of the evolving use of data-matching in Commonwealth administration.

Role of the OAIC

Investigations

16. The OAIC has the function of investigating possible breaches of the Privacy Act, including the IPPs.⁹
17. If an individual considers that an agency or organisation covered by the Privacy Act has interfered with his or her privacy, and they have been unable to resolve the matter directly with the agency or organisation, they can complain to the OAIC. The OAIC may investigate and attempt to resolve the matter by conciliation between the parties.
18. The Privacy Act does not currently impose specific penalties for breaches of the IPPs. However, the Commissioner may make determinations requiring the payment of compensation for damages or other remedies, such as the provision of access or the issuance of an apology.¹⁰ Those determinations can be enforced by the Federal Court or Federal Magistrates Court.¹¹
19. The Commissioner has the power to initiate an investigation on their own initiative in appropriate circumstances, without first receiving a complaint.¹²
20. It is also open to the Commissioner to inform the Minister responsible for the Privacy Act of action that needs to be taken by an agency in order to achieve compliance by the agency with the IPPs.¹³
21. The OAIC conducts its investigations in private. However, where appropriate, the OAIC will publish the details of Commissioner initiated investigations.
22. The Commissioner also has the power to audit agencies covered by the Privacy Act, including agency compliance with the IPPs.¹⁴
23. In some circumstances, consistent with its roles of education and enforcement, the OAIC may publish information about the information management practices of an agency, including audit reports.

Data-matching

24. The Commissioner has a number of specific functions under the Privacy Act with respect to data-matching, including:
 - 24.1. 'to undertake research into, and to monitor developments in, data processing and computer technology (including data-matching and data-linkage) to ensure that any adverse effects of such developments

⁹ Privacy Act, s 27(1)(a).

¹⁰ Privacy Act, s 52.

¹¹ Privacy Act, s 55A.

¹² Privacy Act, s 40(2).

¹³ Privacy Act, s 27(1)(j).

¹⁴ Privacy Act, s 27(1)(h).

on the privacy of individuals are minimised, and to report to the Minister the results of such research and monitoring',¹⁵ and

- 24.2. 'to examine (with or without a request from a Minister or a Norfolk Island Minister) a proposal for data-matching or data linkage that may involve an interference with the privacy of individuals or which may otherwise have any adverse effects on the privacy of individuals and to ensure that any adverse effects of such proposal on the privacy of individuals are minimised'.¹⁶
25. The Privacy Act also enables (and, in some cases, requires) the Commissioner to report to the Minister that administers the Privacy Act about the exercise of specific functions, including the functions set out above.¹⁷

Application of the Guidelines

What scale of data-matching do the Guidelines apply to?

26. Guideline 1 provides that the Guidelines apply to data-matching programs that include the comparison of two or more datasets that, in total, contain information about more than 5000 individuals.

Reporting and notification requirements where the Guidelines do not apply

When should agencies report to the OAIC?

27. Where the proposed data-matching activity involves, in total, the records of fewer than 5000 individuals, but more than 1000 individuals at the end of each financial year, agencies should provide to the OAIC a report on all data-matching programs:
 - 27.1. for which they are the matching agency
 - 27.2. which are not covered by the Guidelines, and
 - 27.3. which involve one or more data-sets that include records relating to, in total, 1000 or more individuals, but fewer than 5000 individuals.
28. On request by the OAIC, each agency should provide to the OAIC a report on all data-matching programs:
 - 28.1. for which they are the matching agency
 - 28.2. which are not covered by the Guidelines, and
 - 28.3. which involve one or more data-sets that include records relating to, in total, 1000 or more individuals, or
 - 28.4. which are not covered by the Guidelines because of Guideline 1.

¹⁵ See Privacy Act, s 27(1)(c).

¹⁶ See Privacy Act, s 27(1)(k).

¹⁷ See Privacy Act, s 32.

Reporting on numerous data-matching programs

29. Reports to the OAIC should list each data-matching program, unless such listing would be impracticable. If an agency conducts numerous similar data-matching programs on an ad hoc basis, and it would not be practicable to list each program individually, the report should include a description of the category of data-matching programs including:
 - 29.1. the purpose of the category of data-matching programs
 - 29.2. what is done with the results of the data-matching programs
 - 29.3. an estimate of the number of data-matching cycles being conducted in the category
 - 29.4. the type(s) of data involved in the programs, and
 - 29.5. what categories of staff within the agency carry out the data-matching programs.

Privacy safeguards

30. If, in the course of participating in a data-matching program not covered by the Guidelines, an agency becomes aware of information that it considers warrants administrative action being taken against an individual, the OAIC considers that it would be good privacy practice to inform the individual and offer them the opportunity to respond regarding the accuracy of the information the subject of the match (in accordance with Guideline 6, below).

Guideline 1 – Application of the Guidelines

1.1 Subject to the exceptions set out in Guideline 1.2, the Guidelines apply to a data-matching program if:

- (a) the data-matching program includes the comparison of two or more datasets that, in total, contain information about more than 5,000 individuals, and
- (b) the information in the datasets was collected for different purposes, and
- (c) the purpose of the data-matching program is:
 - (i) to select individuals for possible administrative action, or
 - (ii) to add information from one database to another for purposes which include taking administrative action in relation to the individuals concerned, or
 - (iii) to add information from one database to another with the intention of analysing the combined information to identify cases where further administrative action may be warranted, or
 - (iv) to permanently combine the databases which provided the datasets being matched by the data-matching program.

1.2 The Guidelines do not apply to a data-matching program if all of the following criteria are satisfied:

- (a) the objective of the data-matching program is to verify personal information collected by the matching agency with another agency or organisation
- (b) the personal information to be verified was provided to the matching agency by the individual or individuals to whom the personal information relates
- (c) the personal information is about the individual's or individuals' circumstances, status or relationship
- (d) the information or results of the data-matching program may be used to determine, or materially contribute to the making of a determination, whether administrative action is warranted, and
- (e) the relevant data-matching cycle of the data-matching program is conducted within 3 months of the collection of the personal information by the matching agency.

Similar data-matching programs to be treated as a single data-matching program

31. If a matching agency runs several very similar data-matching programs (for example, programs which have one data-set and the algorithm in common, but vary as to the other data-set), they should be treated as a single data-matching program for the purpose of assessing whether the Guidelines apply. In

particular, the other source data-sets should be regarded as a single data-set when deciding whether the data-sets used in the data-matching program contain records about more than 5000 individuals.

32. If an agency conducts a number of similar data-matching programs that have the same objective and allow the drawing of similar inferences about the individuals identified, the agency should treat those programs as a single data-matching program for the purpose of complying with these guidelines.

Data-matching programs involving more than one agency

33. Different agencies generally hold information for different purposes. Accordingly, where more than one agency participates in a data-matching program, the Guidelines will likely apply to that program (see Guideline 1.1(b)).
34. However, where two or more agencies have joint responsibility for managing a data-matching program, it is possible that those agencies may each maintain a database of similar or identical information collected for the same purpose. If so, as the information was not collected for different purposes, the Guidelines would be unlikely to apply to the matching of data-sets sourced from those databases.

Guideline 2 – Deciding to carry out or participate in a data-matching program

When deciding to carry out or participate in a data-matching program, or to recommend that such a program should commence, an agency should take into account:

- (a) the costs and benefits for the proposed data-matching program, and
- (b) whether there are any alternative measures to data-matching that could achieve the same results as the proposed data-matching program.

35. There is significant potential for data-matching to be privacy invasive. As such, the OAIC considers that it is best privacy practice to only carry out data-matching where there is a clear business case, having regard to:
- 35.1. the financial and non-financial costs and benefits (see Appendix C for guidance on how to assess the costs and benefits of a data-matching program), and
 - 35.2. whether the desired outcome could practicably be achieved by less privacy invasive means.
36. The OAIC has published a *Privacy Impact Assessment Guide*¹⁸ which provides guidance on, amongst other things, how to assess the privacy impacts of a project. The OAIC encourages agencies that are considering whether to participate in a data-matching project (particularly new or large projects) to carry out a Privacy Impact Assessment of the project.

¹⁸ See www.privacy.gov.au/materials/types/download/9509/6590.

Guideline 3 – Prepare a program protocol

- 3.1 Before commencing a data-matching program, the primary user agency should:
 - (a) prepare a program protocol in accordance with Appendix A
 - (b) provide a copy of the program protocol to the OAIC, and
 - (c) make the program protocol publicly available.
- 3.2 Each agency involved in a data-matching program should ensure that its participation complies with the program protocol.

Purpose of the program protocol

37. The purpose of the program protocol is to inform the public about the existence and nature of the data-matching program.

What should the program protocol contain?

38. The program protocol should provide the following:
 - 38.1. **a description of the program**, including:
 - 38.1.1. an overview of the program
 - 38.1.2. the objectives of the program
 - 38.1.3. the matching and source agencies, and any agencies that will use the results of the program
 - 38.1.4. a description of the data to be provided and the methods used to ensure it is of sufficient quality for use in the program
 - 38.1.5. a brief description of the matching process, the output produced and the destination of the results of the program
 - 38.1.6. what action, administrative or otherwise, may be taken as a result of the program
 - 38.1.7. time limits applying to the conduct of the program, and
 - 38.1.8. what form of notice has been given, or is intended to be given, to individuals whose privacy is affected by the program, and
 - 38.2. **an explanation of the reasons for deciding to conduct the program**, including:
 - 38.2.1. the program's relationship to the agencies' lawful functions and activities
 - 38.2.2. the legal authority for the uses and disclosures of personal information involved in the program
 - 38.2.3. alternative measures to data-matching that were considered, and the reasons why they were discounted
 - 38.2.4. information about any pilot testing of the program, and

38.2.5. a statement of the costs and benefits of the program (see Appendix B for a description of what the statement should contain).

39. A suggested format, and more detailed guidance on the elements of a program protocol, is set out in Appendix A.

Program protocols for similar data-matching programs treated as a single program

40. Where a number of similar data-matching programs have been treated as a single data-matching program under Guideline 1, the program protocol prepared to cover those data-matching programs should deal with the matters set out in Appendix A and should also set out:

40.1. each data source used,

40.2. how many different data-matching programs are involved, and

40.3. what classes of agency staff are responsible for conducting them.

Note: Where a data-matching program is current at the date of commencement of these Guidelines and a program protocol has not been prepared, the primary user agency should ensure that, within a timeframe agreed with the OAIC, a protocol is prepared, forwarded to the OAIC, and made publicly available.

Publishing the program protocol

41. With respect to Guideline 3.1(c), the OAIC recommends that the primary user agency publish the program protocol for the data-matching program on its website.

42. Agencies should also consider their obligations under the FOI Act, particularly in relation to the Information Publication Scheme (IPS). Under the IPS, agencies are required to publish information that falls within the categories specified in s 8(2) of the FOI Act. In particular, agencies should consider whether the program protocol for a data-matching program falls within the scope of the agency's 'operational information',¹⁹ and should therefore be published as part of the IPS.

43. Further guidance on the IPS is set out in Part 13 of the *Guidelines issued by the Australian Information Commissioner under s 93A of the Freedom of Information Act 1982*.²⁰

¹⁹ See FOI Act, s 8(2)(j), s 8A.

²⁰ See www.oaic.gov.au/publications/guidelines/part13_ips.html.

Changes to data-matching programs

44. When a matching agency wishes to change or amend an existing data-matching program, the matching agency should revise the program protocol to clearly indicate the amendments.
45. The matching agency should notify the OAIC of the amendments, and provide the OAIC with a copy of the revised program protocol.
46. The amendments should be considered in the program evaluation (see Guideline 9).

Guideline 4 – Prepare a technical standards report

- 4.1 Before commencing a data-matching program, the matching agency should prepare detailed technical standards to govern the conduct of the data-matching program.
- 4.2 Where practical, the technical standards should be developed in consultation with source agencies or organisations.
- 4.3 The matching agency should detail the technical standards in a technical standard report that includes the matters set out in Appendix B.

Purpose of the technical standards report

47. IPP 4 requires that agencies ensure that the personal information they hold is protected by reasonable security safeguards to ensure that the information is protected against loss, unauthorised use, modification or disclosure, and against other misuse. Further, IPP 7 and IPP 8 require that agencies ensure the accuracy of the personal information they hold.
48. Accordingly, the OAIC considers that it is best practice to have clearly expressed and detailed technical standards.
49. The Attorney-General's Department has published the *Improving the Integrity of Identity Data – Data Matching Better Practice Guidelines*,²¹ which provides guidance that may assist in establishing appropriate technical standards.

What should the technical standards report contain?

50. The technical standards report should include the following:
 - 50.1. **a description of data** supplied by source agencies or organisations
 - 50.2. **the specification** for each matching algorithm or project,
 - 50.3. **any risks** inherent in the data-matching program, and how those risks will be addressed,
 - 50.4. **controls** to be employed to ensure the continued integrity of the data used in the data-matching program, and of the data-matching program as a whole, and
 - 50.5. **security features** included in the program to control and minimise access to personal information.

²¹ See

[www.ag.gov.au/www/agd/rwpattach.nsf/VAP/\(8AB0BDE05570AAD0EF9C283AA8F533E3\)~Data+Matching+Guidelines+updated+15102010+RC.pdf/\\$file/Data+Matching+Guidelines+updated+15102010+RC.pdf](http://www.ag.gov.au/www/agd/rwpattach.nsf/VAP/(8AB0BDE05570AAD0EF9C283AA8F533E3)~Data+Matching+Guidelines+updated+15102010+RC.pdf/$file/Data+Matching+Guidelines+updated+15102010+RC.pdf).

51. More detailed guidance on the elements of a technical standards report is set out in Appendix B.

Comply with the technical standards report

52. The technical standards report should be prepared and held by the matching agency, and copies held by the source and user agencies or organisations where this is practicable.
53. Each agency participating in a data-matching program (in any capacity) should ensure that its participation in the program is in accordance with the technical standards report.
54. The matching agency should provide a copy of the technical standards report to the OAIC on request. The technical standards report will assist in the proper assessment of an agency's compliance with the IPPs, in that it presents information in a manner that is capable of independent scrutiny. Accordingly, the technical standards report may be used as a basis for any review of the actual data-matching activity that the OAIC may conduct (see Guideline 12).

New data-matching programs

55. For new data-matching programs, the technical standards report should be completed in draft form prior to the commencement of the program. It should be finalised not later than 30 days after the end of the first cycle, taking account of the initial experience of the operation of the program.

Changes to data-matching program specifications

56. When a matching agency wishes to amend the specifications of an existing data-matching program, the matching agency should revise the technical standards report to clearly indicate the amendments.

Note: Where a data-matching program is current at the date of commencement of these guidelines and a technical standards report has not been prepared, the matching agency should ensure that it prepares one within a timeframe agreed with the OAIC.

Guideline 5 – Notify the public

Before an agency carries out or participates in a data-matching program, the agency should take reasonable steps to ensure public notice of the proposed program is given.

Obligation to notify

57. IPP 2 requires agencies that collect personal information about an individual to take reasonable steps, before the collection or as soon as practicable after, to notify the individual of certain matters including:

- 57.1. why the agency is collecting the information
- 57.2. whether the agency has legal authority to collect the information, and
- 57.3. to whom the agency usually discloses that kind of information.

Content of public notice

58. Public notification of a data-matching program should, as a minimum:

- 58.1. contain a brief description of the objectives of the data-matching program
- 58.2. list the agencies involved in the data-matching program
- 58.3. list the categories of information contained in the data sets involved in the data-matching program
- 58.4. list the categories of individuals about whom personal information is to be matched, and
- 58.5. include the approximate number of individuals affected.

Forms of public notice

Primary user agency

- 59. The primary user agency for a proposed data-matching program should cause a notice to be published in the Commonwealth Government Gazette.
- 60. The Gazette notice should be published before the commencement of the data-matching program.
- 61. The primary user agency should forward copies of the Gazette notice to any individuals, agencies or organisations nominated by the OAIC (if any).

All participating agencies

- 62. Each agency or source data organisation participating in a data-matching program in any capacity should take reasonable steps to notify the general public and affected individuals about the data-matching program.

63. In addition to a notice in the Commonwealth Government Gazette, agencies could notify the general public by, for example:
 - 63.1. including a notification on the agency's website, or
 - 63.2. placing advertisements in print or online media publications.
64. Agencies or source data organisations should also take steps specifically aimed at informing individuals whose information is likely to be used in the program. This could be done by, for example:
 - 64.1. including information about the proposed data-matching program in material given to individuals when they provide information that is likely to be used in the data-matching program
 - 64.2. informing relevant clients about the proposed data-matching program directly (for example, by letter or email), or
 - 64.3. by placing notices in relevant special-purpose publications or newsletters.
65. Where a number of similar programs have been treated as a single program (see Guideline 1), the Gazette notice and other publicity material should deal with the matters set out above (see 'Content of public notice' above) and should also describe:
 - 65.1. the range of data sources used
 - 65.2. how many different data-matching programs are involved, and
 - 65.3. what classes of agency staff are responsible for conducting those data-matching programs.
66. The Gazette notice and other publicity material should advise how the general public can obtain copies of the program protocol.

Note: Where a data-matching program is current at the date of commencement of these Guidelines, and public notice of the program has not been given, the primary user agency should ensure that, within a timeframe agreed with the OAIC, public notice of the program is given in accordance with Guideline 5.

Submission for inclusion in the Personal Information Digest

67. IPP 5(3) requires agencies to maintain a record setting out:
 - 67.1. the nature of the various types of records of personal information kept by the agency
 - 67.2. the purpose for which the records are kept
 - 67.3. the class of individuals to which the records apply
 - 67.4. the period for which the records are kept
 - 67.5. details of how individuals can get access to records about themselves.

68. That record must be made available for public inspection on request, and be provided to the OAIC at the end of each financial year.²²
69. The OAIC compiles those records and publishes them annually as the Personal Information Digest (PID).²³
70. A matching agency should include a description of any records of personal information it holds in connection with a data-matching program in its records of its personal information holdings, and in its submission for inclusion in the PID.
71. A source agency should include a note that the records are disclosed to the matching agency in connection with a data-matching program, in the relevant part of its PID submission.

²² Privacy Act, s 14, IPP5(4).

²³ Privacy Act, s 27(1)(g).

Guideline 6 – Notify individuals of proposed administrative action

Before taking administrative action against an individual in response to a match, agencies should notify the individual of the match and the proposed administrative action, and give the individual the opportunity to respond.

72. IPP 8 requires that agencies ensure the accuracy of the personal information they hold before using that information. IPP 7 also requires agencies to take steps to amend records that contain personal information to ensure that personal information is accurate.
73. In relation to a match produced as the result of a data-matching program, methods to ensure the accuracy of the match could include checking the data with third parties, against source data, or with the individual the subject of the match.
74. As a matter of best privacy practice, a user agency should only take administrative action in response to a match after giving the individual concerned:
 - 74.1. reasonable notice of the relevant matters including:
 - the match
 - the conclusions the agency has drawn based on the match, and
 - the administrative action that the agency proposes to take in response to the match, and
 - 74.2. a reasonable period in which to respond to that information.
75. In the view of the OAIC, individuals should be given at least **14 days** to respond.
76. If there is a dispute as to the accuracy of the data, but the agency considers that administrative action is still warranted, it should inform the individual of their right to lodge a complaint with the OAIC regarding the accuracy of their personal information.
77. If an agency plans to take administrative action in response to a match without notifying the individual concerned of the match (for example, as part of a data-matching program that does not comply with the Guidelines; see Guideline 10), the agency should take reasonable steps to ensure the accuracy of the information before taking action.
78. The decision to take administrative action should ideally involve consideration of the circumstances of each case. However, where a program is generating large numbers of matches, it may be reasonable to apply some rule based selection criteria, to achieve further filtering or selection of cases independently of the matching program itself. The filtering criteria should be described in the technical standards report (see Guideline 4).

79. Unless required or authorised by law, an agency should not take administrative action that interferes with the individual's opportunity to exercise any rights of appeal or review.

Guideline 7 – Destroy information that is no longer required

At the conclusion of a data-matching program, agencies should destroy or de-identify personal information that is no longer required.

80. IPP 4 requires that agencies ensure that the personal information they hold is protected by reasonable security safeguards to ensure that the information is protected against loss, unauthorised use, modification or disclosure, and against other misuse. In the view of the OAIC, destruction or de-identification of personal information that is no longer required will usually be a reasonable step to prevent the loss or misuse of that information.
81. Destruction of records collected for the purpose of data-matching should be conducted in accordance with the National Archives of Australia's *General Disposal Authority 24 – Records Relating to Data Matching Exercises*.²⁴
82. That document provides that:
 - 82.1. where personal information is obtained for use in a 'data-matching exercise' (ie, a data-matching cycle as part of a data-matching program), and
 - 82.2. the 'data matching process' (ie, the data-matching cycle involving the information) does not lead to a match,then the personal information should be destroyed by the matching agency, where practicable, within 14 days of the completion of the data-matching cycle, and at least within 90 days of the completion of the data-matching cycle.
83. The Disposal Authority provides that the Commissioner may approve an extension of time for the destruction of such information. Agencies that seek such approval must make a written request to the OAIC. The request should be made in accordance with Guideline 10.
84. In relation to matches that may or may not result in administrative action, agencies should carefully consider their obligations under IPP 4 with respect to the retention and destruction of that information.

²⁴ See www.naa.gov.au/Images/GDA24_tcm2-1128.pdf.

Guideline 8 – Do not create new registers, data-sets, or databases

An agency involved in a data-matching program, in any capacity, should not create any new separate permanent register, data-set, or database using data-sets or information contained in data-sets collected as part of the data-matching program.

85. IPP 10 requires that agencies must not use personal information for a purpose other than the purpose for which it was collected, unless a listed exception applies.
86. Privacy concerns relating to data-matching include the possibility that agencies will:
 - 86.1. use personal information collected for the purpose of a specific data-matching program for an unrelated secondary purpose, or
 - 86.2. retain personal information indefinitely in case it becomes useful in future.
87. Unless a secondary use is authorised by an exception listed in IPP 10, personal information collected for the purpose of data-matching should be destroyed when no longer required (see Guideline 7)
88. Compliance with Guideline 8 would not preclude:
 - 88.1. the maintenance of a register of individuals in respect of whom further investigations are warranted under the terms of the program protocol, following a decision to take administrative action involving those individuals
 - 88.2. the maintenance of a special register solely for the purpose of excluding individuals from being selected for investigation in successive data-matching cycles of the same data-matching program, or
 - 88.3. the maintenance of such records or databases as are reasonably necessary to achieve the objectives of the program.

Guideline 9 – Regularly evaluate data-matching programs

The primary user agency should evaluate the conduct and outcomes of data-matching programs no later than three years after the commencement of operation of the data-matching program, and at least every three years after that while the program continues.

89. Prior to participating in a data-matching program, agencies should ensure that their decision to participate is based on a sound business case, and is in the public interest having regard to the potential for data-matching to be privacy invasive (see Guidelines 2 and 3).
90. However, where data-matching programs are conducted over the course of several years, it is important for agencies to periodically confirm that the reasons for participating in the data-matching program are still valid – for example, that the program is achieving its objectives and has not deviated from the privacy and data quality safeguards specified in the program protocol and technical standard report.
91. Accordingly, no less than every three years, the primary user agency should undertake an evaluation of the data-matching program in accordance with its original objectives.
92. Where the primary user agency is not the matching agency, the primary user agency should consult with the matching agency regarding the evaluation.
93. The evaluation should include (as a minimum):
 - 93.1. consideration of whether the data-matching program has achieved its objectives
 - 93.2. consideration of whether the program has complied with the program protocols and technical standards report
 - 93.3. consideration of whether the privacy and data quality safeguards incorporated into the data-matching program have been effective
 - 93.4. a revised statement of the costs and benefits of the program (see Appendix C)
 - 93.5. a determination as to whether the reasons for conducting the data-matching program are still valid, and details of any changes or amendments to the data-matching program during the evaluation period, or as a consequence of the evaluation.

94. As a matter of best practice, the primary user agency should:
 - 94.1. document the conclusions of the evaluation in a report
 - 94.2. make a report of the evaluation publically available (by, for example, posting a copy on the agency website)
 - 94.3. provide a copy of the report to the OAIC.

Note: Where the data-matching program has not been publically notified (for example, where the data-matching program relates to confidential information sourced from law enforcement or national security agencies) it may not be appropriate to make a copy of the evaluation report publically available.

Guideline 10 – Seeking exemptions from Guideline requirements

Where the head of an agency considers that it would be appropriate (having regard to the public interest) to conduct a data-matching program to which the Guidelines apply in a way that would be inconsistent with one or more of the Guidelines, they should:

- (a) advise the OAIC in writing of the details of the proposed data-matching program
- (b) in that advice, specify how the proposed data-matching program would be inconsistent with the Guidelines, and
- (c) explain the public interest grounds that justify the inconsistency.

Explaining the public interest grounds

95. In explaining the public interest grounds, the head of the agency should address matters such as the following:
 - 95.1. the effect that not abiding by the Guidelines would have on individual privacy
 - 95.2. the seriousness of the administrative or enforcement action that may flow from a match obtained through the data-matching program
 - 95.3. the effect that not abiding by the Guidelines would have on the fairness of the data-matching program – including its effect on the ability of individuals to determine the basis of decisions that affect them, and their ability to dispute those decisions
 - 95.4. the effect that not abiding by the Guidelines would have on the transparency and accountability of agency and Government operations
 - 95.5. the effect that not abiding by the Guidelines would have on compliance of the proposed data-matching program with the IPPs
 - 95.6. the effect that complying with the Guidelines would have on the effectiveness of the proposed data-matching program
 - 95.7. whether complying fully with the Guidelines could jeopardise or endanger the life or physical safety of information providers or could compromise the source of information provided in confidence
 - 95.8. the effect that complying fully with the Guidelines would have on public revenue – including tax revenue, personal benefit payments, debts to the Commonwealth and fraud against the Commonwealth
 - 95.9. whether complying fully with the Guidelines would involve the release of a document that would be an exempt document under the FOI Act, and
 - 95.10. any legal authority for, or any legal obligation that requires, the conduct of the proposed data-matching program in a way that is inconsistent with the Guidelines.

OAIC response to advice

96. The Commissioner may respond to the agency head's advice, setting out his or her view as to whether it would be appropriate, from a privacy protection perspective, for the Guidelines not to be followed, and the reasons for taking this view.
97. If the Commissioner takes the view that it would be inappropriate, from a privacy protection perspective, for the Guidelines not to be followed, the Commissioner may suggest changes to the proposed data-matching program which would, in the Commissioner's view, achieve an adequate standard of privacy protection.
98. The Commissioner cannot exempt an agency from the requirements of the IPPs in the absence of a relevant Public Interest Determination.²⁵

Publication of advice

99. It is the normal practice of the OAIC to make an advice provided in accordance with Guideline 10 publically available. However, the OAIC will keep such advice confidential if, in the advice, the agency head:
 - 99.1. requests that the advice remain confidential, and
 - 99.2. provides adequate reasons for that request.
100. Freedom of Information requests for advice provided on a confidential basis will be considered in accordance with the FOI Act.

²⁵ See Privacy Act, Part VI.

Guideline 11 – Data-matching with entities other than agencies

Where an agency proposes to carry out a data-matching program that involves an entity that is not an agency (such as a State or Territory government body, or a private sector organisation), whether as the matching agency or a source agency, the agency should require that the entity adopt these Guidelines in respect of the data-matching program.

102. Under s 95B of the Privacy Act, agencies seeking to carry out a data-matching activity with an organisation (under contract) must take contractual measures to ensure that the organisation carries out the contract work in a manner consistent with the IPPs.
103. A 'contractual measure' in this context might include, where practicable, requiring non-agency participants in a data-matching program to adopt these Guidelines.
104. For example, where an agency enters into a contractual agreement to carry out data-matching with the participation of a non-agency party, the agency could seek that the contract includes a condition requiring the parties to the contract to adopt these Guidelines.

Guideline 12 – Enable review by the OAIC

Agencies should enable the OAIC to review their data-matching activities and their procedures.

105. The review process carried out by the OAIC may include:
 - 105.1. assessing whether the data-matching program is being conducted in accordance with the procedures set out in the program protocol,
 - 105.2. reviewing the effectiveness of the controls and procedures set out in the technical standards report
 - 105.3. assessing the outcomes of the data-matching program from a privacy perspective, and
 - 105.4. considering any complaints and difficulties that have arisen in connection with the data-matching program.
106. On request from the OAIC, agencies should report to the OAIC on any relevant matter, including:
 - 106.1. actual costs and benefits flowing from the data-matching program
 - 106.2. any non-financial but quantifiable factors that are considered relevant
 - 106.3. any difficulties in the operation of the program and how these have been overcome
 - 106.4. the extent to which internal audits or other forms of assessment have been undertaken by the participating agencies or organisations, and their outcome; and
 - 106.5. the number of matches, the number of matches investigated, the number of cases not proceeded with after contacting the affected individual, and the number of cases in which action proceeded despite a challenge as to accuracy of the data.
107. It is the usual practice of the OAIC to include in its annual report general information about:
 - 107.1. the number, extent and nature of data-matching programs
 - 107.2. the extent of public notification of programs and of consultation
 - 107.3. the extent of confidential notification of programs to the OAIC
 - 107.4. the nature of the public interest reasons advanced for not engaging in public notification of programs; and
 - 107.5. the operational experience and effectiveness of programs.

Appendix A: Content of data-matching program protocols

The purpose of the program protocol is to inform the public about the existence and nature of the data-matching program.

Accordingly, the program protocol should be written in plain English, and give an accurate picture of how the data-matching program works.

The program protocol should provide the following:

- **a description of the data-matching program**, including:
 - (a) an overview of the data-matching program
 - (b) the objectives of the data-matching program
 - (c) the matching and source agencies, and any agencies that will use the results of the data-matching program
 - (d) a description of the data to be provided and the methods used to ensure it is of sufficient quality for use in the data-matching program
 - (e) a brief description of the matching process, the output produced and the destination of the results of the data-matching program
 - (f) what action, administrative or otherwise, may be taken as a result of the data-matching program
 - (g) time limits applying to the conduct of the data-matching program, and
 - (h) what form of notice has been given, or is intended to be given, to individuals whose privacy is affected by the data-matching program; and
- **an explanation of the reasons for deciding to conduct the data-matching program**, including:
 - (i) the data-matching program's relationship to the agencies' lawful functions and activities
 - (j) the legal authority for the uses and disclosures of personal information involved in the data-matching program
 - (k) alternative measures to data-matching that were considered, and the reasons why they were rejected
 - (l) information about any pilot testing of the data-matching program, and
 - (m) a statement of the costs and benefits of the data-matching program (see Appendix B for a description of what the statement should contain).

A suggested format, and some guidance on the elements of a program protocol, is set out below.

Description of the data-matching program

The description of the data-matching program needs to cover the following matters. Suggested section titles are in parentheses.

(a) An overview of the data-matching program (Overview)

A short, simply expressed statement of what the data-matching program does and why. 200–300 words should be sufficient.

(b) The objectives of the data-matching program (Objectives)

A basic statement of what the data-matching program is trying to achieve.

The rest of the protocol will flesh out exactly how the data-matching program will try to achieve its objectives, so this statement does not need to be lengthy.

(c) The matching and source agencies, and any agencies which will use the results of the data-matching program (Agencies involved)

This should include:

- which agency is conducting the data-matching program
- where the matching agency is not the same as the primary user agency, the protocol should clearly establish the functions and roles of each participating agency with respect to the data-matching program
- which agencies or other sources are providing data that will be used in the data-matching program – this should cover all sources of data, including non-Commonwealth and non-government sources, and
- all agencies or other organisations that have access to the results of the data-matching program.

(d) A description of the data to be provided and the methods used to ensure it is of sufficient quality for use in the data-matching program (Data issues)

For each data or information source involved in the data-matching program, briefly describe:

- the kind of files transferred to the matching agency
- the type of information contained in the file
- the approximate number of records on each file, and
- what measures have been taken to ensure the quality, integrity and security of the data.

(e) A brief description of the matching process, the output produced and the destination of the results of the data-matching program (The matching process)

Describe:

- which fields are matched (eg, agency identifier , name and date of birth)
- what criteria are used to identify a 'match' (eg, individuals on both files, individuals on one but not the other), and
- the fields included in each output file.

The specific technical details of the matching process will be set out in the technical standards report, so the description here can be made in relatively broad terms.

(f) What action, administrative or otherwise, may be taken as a result of the data-matching program (Action resulting from the program)

This should cover all agencies that use the results of the matching. If an agency may take one of a range of actions, depending on the facts of a particular case, each should be outlined. Copies of template letters to people that are proposed to be, or will be, the subject of administrative action should be attached to the program protocol as an appendix.

(g) Time limits applying to the conduct of the data-matching program (Time limits)

This should cover:

- how long data obtained for use in the data-matching program will be kept, including both input data provided by source agencies and the output data from the matching;
- retention periods and disposal arrangements for all data; and
- how frequently the program will be run; and, if the data-matching program will be run at infrequent intervals, how it is decided that a run is appropriate at a particular time; and
- when it is planned to terminate or review the data-matching program, including agencies' internal review mechanisms as well as external mechanisms, such as legislative sunset clauses.

(h) What form of notice has been given, or is intended to be given, to individuals whose privacy is affected by the data-matching program (Public notice of the program)

The text of public notices should be attached to the protocol, including the text of IPP 2 notices, gazette notices, media releases and so on.

Reasons for deciding to conduct the data-matching program

The reasons for deciding to conduct the data-matching program should cover the following matters.

(i) The data-matching program's relationship to the agencies' lawful functions and activities. (Relationship to lawful functions)

This should clearly specify the link between the objectives of the data-matching program and each agency's lawful functions and activities.

(j) The legal authority for the uses and disclosure of personal information involved in the data-matching program (Legal authority)

The reasons should include the justification of the use and disclosure of personal information in terms of the Privacy Act and any other relevant legislation.

To be lawful, any use of personal information by an agency for a purpose other than the purpose for which the information was collected must fall within one of the five exceptions listed in IPP 10.

Similarly, any disclosure of personal information by an agency must fall within one of the five exceptions listed in IPP 11.

The protocol should specify which exceptions in IPP 10 and 11 apply, and why.

If there are other legislative requirements regarding the use and disclosure of information involved in the data-matching program, such as secrecy provisions, the protocol should explain how the uses and disclosures are authorised in terms of the relevant provisions.

(k) Alternative measures to data-matching that were considered, and the reasons why they were rejected. (Alternative methods)

If it is considered that there are no practicable alternatives to data-matching, the protocol should include a brief explanation of why this is the case.

(l) Information about any pilot testing of the program. (Pilot programs)

Where a pilot program has been conducted, the following information is likely to be relevant and should be included:

- the number of records involved in the pilot program
- the number of matches that resulted
- an estimate or report of the benefits of the pilot program (if the matches were acted upon, it may be possible to give a detailed account of the benefits; if the matches were not acted upon, an estimate of the benefits that would have resulted should be given), and
- information about any problems or difficulties with the matching program that was obtained from the pilot program.

If the protocol relates to a new data-matching program (rather than one already operating) and no pilot project has been conducted or is planned, the protocol should indicate why a pilot program is not considered to be necessary.

(m) A statement of the costs and benefits of the data-matching program. (Costs and benefits)

See Appendix C regarding the preparation of a statement of costs and benefits.

Appendix B: Technical standards report

Purpose

The purpose of the technical standards report is:

- to ensure that data-matching is conducted on the basis of pre-defined standards, including data-quality and security controls, and
- to form a basis for any review of the actual data-matching activity that the OAIC may conduct.

Contents

The technical standards report should include the following:

- (a) **a description of data** supplied by source agencies or organisations, referring in particular to:
 - key terms and definitions
 - the scope and completeness of the data to be collected
 - the relevance of the data to be collected, and
 - the timing of the collection,
- (b) **the specification** for each matching algorithm or project, including such things as:
 - data items used in the match, particularly the use of any Government identifiers,
 - the rules for recognising a match, and
 - the destination of the results of the data-matching program,
- (c) **any risks** inherent in the data-matching program, and how those risks will be addressed,
- (d) **controls** to be employed to ensure the continued integrity of the data used in the data-matching program, and of the data-matching program as a whole, and
- (e) **security features** included in the program to control and minimise access to personal information.

(a) Description of data

As part of the technical standards report, the matching agency should, in consultation with the source agencies or organisations where that is practicable, compile a 'data dictionary' for all data that is supplied as part of the data-matching program that includes the following:

- a description of each file used by the matching agency that outlines its source, destination and, for an intermediate file, its use
- for each data item:
 - its name, description, the validation or edits applied to it

- whether or not it has been standardised, and
- the level of precision of the field, for example, YY or YY/MM/DD, annual income, amount in thousands.

(b) Matching techniques

The technical standards report should clearly document the following information about the data-matching techniques to be used in the data-matching program:

- the matching algorithm used – for example, first 6 characters of surname and value of forename, together with date of birth; phonic equivalent of family name and equality of birth year
- rules for recognising matches
- the destination of matching results
- the sampling techniques used to verify the validity/accuracy of matches, and
- the techniques adopted to overcome identifiable problems with the quality of data and to standardise data items that have been compared but have different meaning (for example, 'annual income' and 'financial year income').

(c) Risks

The technical standards report should identify any risks posed by the data-matching program including, but not limited to, risks to the privacy of individuals, reputational risks, and risks relating to incorrect matches.

The technical standards report should clearly set out how those risks will be mitigated. Some risks may be mitigated through data quality and security controls (see below at (d) and (e)).

(d) Data quality controls and audit

The technical standards report should clearly document the following:

- any relevant measures taken to ensure data quality (for example, the timing of any extract files that may be taken for the data-matching program), and
- any audit processes to which the data used in the data-matching program has been, or is regularly, subjected.

(e) Security and confidentiality

The technical standards report should clearly document the precautions proposed be taken at all stages of a data-matching program to ensure that personal information used in and arising from a data-matching program:

- is not subject to accidental or intentional modification
- is not accessed by staff within the agency except where such access is necessary for the conduct of that data-matching program or resulting action, and
- is not disclosed otherwise than as intended by the program protocol.

The technical standards report should make specific reference to access controls such as password security, encryption, and audit trails including logging of access.

The technical standards report should also contain a list of all computer programs developed by the matching agency in relation to the data-matching activity, together with a description of the functions of the programs.

Appendix C: Statement of costs and benefits for data-matching programs

Introduction

Guideline 4 calls for agencies to prepare statements of costs and benefits:

- when starting a new data-matching program (as part of the program protocol)
- when preparing a program protocol for a data-matching program that was already running when the matching agency adopted the Guidelines, and
- when evaluating data-matching programs.

While it is desirable for the statement to be as comprehensive and rigorous as possible, it is not intended to be a formal cost-benefit analysis. In some cases the information required for such an analysis will not be available. In others, the sort of net benefit bottom line that a formal cost-benefit analysis aims to produce will not be the most meaningful way of presenting the impact of the data-matching program.

The appropriate degree of detail will vary depending on the nature of the data-matching program. For example, a less detailed statement might be appropriate for a data-matching program whose function is to carry out a task that would otherwise have to be performed manually.

The costs and benefits typically associated with data-matching programs are categorised below, together with suggestions on how to present information in each category. The categories can be used as a check-list by agencies. However, it is important to note that not all of the categories of costs and benefits below will apply to every data-matching program.

Purpose of estimating costs and benefits

The purpose of including a statement of costs and benefits in program protocols and program evaluations is:

- to explain the agency's reasons for determining that the data-matching program is in the public interest
- to help identify areas of potential risk (such as cost, legal liability or public sensitivity), and
- to provide a basis for evaluating the performance of the data-matching program.

A statement that gives aggregate figures for costs and benefits but does not explain how they were calculated is not informative to readers, and does not give a good basis for comparison with actual performance.

Where a statement provides estimated costs and benefits, it is critical that the statement clearly specifies the method by which the estimates were reached, and any assumptions on which they were based.

Methods of presenting cost/benefit information

Perspective

The OAIC suggests that statements of costs and benefits be presented from the perspective of the Commonwealth, rather than from the perspective of the individual agency or the wider community. This means that statements should present information about all significant costs and benefits to the Commonwealth, including costs and benefits experienced by more than one agency.

If a data-matching program will have major costs or benefits for other parts of the community (sometimes called 'externalities'), this should be noted in the statement.

Sources of information

Key sources of data on costs and benefits will be:

- for data-matching programs that are already running, data obtained from the actual operation of the program. This should include (in addition to more detailed information on costs and benefits) basic data on:
 - the total number of matches
 - the number of cases in which matches led to further investigation, and
 - the outcomes from investigation of cases, and
- for new data-matching programs, any pilot program or other preliminary assessment of the data-matching program. If estimates of costs and benefits are based on results from pilot program, those results should be included.

While international comparisons may be useful in limited circumstances, the grounds of comparison are rarely firm. Countries inevitably differ in numerous ways – cultural, economic, legal and so forth – that make parallels difficult to sustain. Such comparisons should be used with caution, if at all, and should not play a pivotal role in arriving at estimates of costs and benefits.

Methods of presenting information

The best way of presenting the information will depend on the nature of the data-matching program and the information available. However, as a minimum, the statement should compare the outcomes from the data-matching program and the outcomes that would arise in the absence of the data-matching program.

Agencies could consider the following formats for statements of costs and benefits.

- Compare the costs and benefits of the data-matching program with the most likely alternative use of the resources required, that is, what those resources would be used for if the program did not go ahead. The benefits of the

alternative use of resources are, effectively, the cost of carrying out the program (as they represent the opportunity cost of devoting those resources to the program). For example, if the resources used in the program have been diverted from a random audit program, then that would be the appropriate basis for comparison.

- Compare the costs and benefits of conducting the data-matching program against the costs and benefits of achieving the same result by some alternative method (for example, a manual process). This may be most appropriate where:
 - one of the main benefits of the data-matching program is administrative savings from the more efficient performance of a task that would otherwise have to be carried out by other means, or
 - the benefits of the data-matching program are quantifiable, but are not financial (for example, detection and prosecution of people breaking a law).
- If the alternative uses of the resources required for the data-matching program, or the costs of achieving the same outcomes by alternative means, cannot be ascertained, estimate the actual costs of the resources to be used in the program, and the expected benefits from the program, without making a comparison with an alternative scenario. This option should only be used where it is impossible to apply either of the other approaches; it gives much less useful information about the data-matching program and makes it much more difficult to judge whether it constitutes an efficient application of resources. If this approach is taken, the statement should include the reasons why neither of the other approaches could be taken.

Calculating net present value

Formal techniques for cost/benefit analysis often include provision for 'discounting' of costs and benefits that occur later, to arrive at a net present value for a data-matching program.

This type of calculation is only necessary for data-matching programs that have a high establishment cost (for example, if they involve significant capital expenditure) and long term benefits.

If costs and benefits are discounted, the undiscounted values should still be presented in the statement, and the discount rate explicitly stated.

Information on calculating net present values can be found on page 18 of the Department of Finance and Deregulation publication *Introduction to Cost-Benefit Analysis and Alternative Evaluation Methodologies* (January 2006).²⁶

²⁶ See www.finance.gov.au/publications/finance-circulars/2006/docs/Intro_to_CB_analysis.pdf.

Further information on cost/benefit analysis

The Department of Finance and Deregulation has produced the following publications relevant to cost/benefit assessment:

- *Guidelines for Costing Government Activities* (1991)²⁷
- *Introduction to Cost-Benefit Analysis and Alternative Evaluation Methodologies* (January 2006), and
- *Handbook of Cost-Benefit Analysis* (January 2006).²⁸

Estimating costs

The costs of data-matching programs can be broadly divided into:

- (a) **establishment costs**, comprising:
- (i) staff costs involved in setting up the data-matching program, for example, staff time to develop appropriate systems to process and handle the results
 - (ii) capital costs, and
 - (iii) other costs, such as publicity costs and computer time, and
- (b) **running costs**, including:
- (i) costs associated with conducting runs of the data-matching program, for example the cost of maintaining the system, and
 - (ii) costs associated with taking action in response to matches, such as administrative costs associated with corresponding with affected individuals or amending benefits, legal costs, or the costs of appeals processes.

For many data-matching programs, the costs associated with some of these categories may be negligible (for example, many data-matching programs do not involve capital expenditure) and can be ignored.

(a) Establishment costs

Staff costs

The OAIC suggests the following approach to estimating the cost of staff time involved in developing a new data-matching project:

Step 1: Estimate the amount of time (in person weeks, months or years) that will be spent by all staff on initial development of the project.

Step 2: Estimate the average salary of project staff (per week, month or year), and multiply it by the estimated staff time needed for project development.

²⁷ Available from the Australian Government Publishing Service.

²⁸ See www.finance.gov.au/publications/finance-circulars/2006/docs/Handbook_of_CB_analysis.pdf.

Step 3: Multiply the result of step 2 by a factor to allow for labour on-costs and overheads. The Department of Finance and Deregulation recommends multiplying the basic salary cost by a factor of 2.54 to get the total staff costs (see Chapter 5 of the Department of Finance 'Guidelines for the Costing of Government Activities').

The staff costs must include time spent on the project by all staff, including administrative, corporate services, IT and support staff, as well as staff dedicated to the project. It is not acceptable to minimise the apparent costs of a data-matching program by trying to shift costs associated with the program to other areas of the agency.

The statement of costs and benefits should indicate the amount of staff time estimated for project development, the total staff cost of setting up the project and the method of calculation.

Capital costs

Most data-matching programs do not require capital expenditure. If capital outlays are required for a program (for example, if computer facilities are to be expanded to cater for the project) they should be stated (see page 13 of the Department of Finance and Deregulation '*Guidelines for Costing of Government Activities*' for more information on assessing capital costs).

Other costs

It is not necessary to include other costs, such as publicity costs and use of computer facilities unless they are significant in magnitude. As a general rule, it will not be necessary to quantify such costs if they represent only a few per cent of overall establishment costs.

Agencies should ensure that all other costs, including other costs incurred by IT, corporate services or special projects areas are considered in this category.

(b) *Running costs*

Cost of conducting matching

The cost of conducting the data-matching program would include:

- staff time (of both IT staff and administrative staff with continuing responsibility for managing the program), and
- computer time.

If these costs are negligible in size they may be excluded from the statement of costs; a statement that the costs are negligible and have been excluded should be included in the protocol.

Costs associated with responding to matches

The costs associated with taking action in response to matches obtained through a data-matching program will also tend to be predominantly comprised of staff costs.

The method used to estimate those costs will depend on the way the data-matching program results are used.

- If the response to matches is carried out by dedicated staff, the cost of this activity can be readily calculated.
- If staff who carry out reviews based on matches also have other functions, the time required for those reviews could be calculated either by estimating the proportion of time that the review staff spend on this activity, or by estimating the time required for an average review, and multiplying it by the number of reviews undertaken. The latter approach would be most suitable in situations where the task of responding to matches is decentralised, and data on how review staff allocate their time is not available.

Estimating benefits

The benefits of data-matching programs can be broadly divided into:

- (a) **direct financial benefits**, including
 - recovery of incorrect payments
 - prevention of incorrect payments, and
 - increased revenue collection.
- (b) **indirect financial benefits**, such as
 - administrative savings
 - benefits of voluntary compliance (deterrence effects), and
- (c) **non-financial benefits**.

Direct financial benefits

Direct financial benefits will mainly fall into the following three categories.

Recovery of incorrect payments

The total amount of overpayments identified as a result of the program should be reduced to recognise:

- that not all overpayments identified will be recovered, because some amounts are too small to warrant recovery action, and in some cases because recovery action will be unsuccessful, and
- the cost of recovery action (unless this is explicitly included under the costs of the program).

For example, an agency may estimate that 70 per cent of identified overpayments represent actual savings, i.e., the remaining 30 per cent of overpayments will be unable to be recovered, or the recovery costs would exceed the amount to be recovered.

The statement of benefits should include both the total amount of overpayments identified and the method of calculating how much of those overpayments represent actual savings.

If full figures for recovered amounts and the costs of recovery are available (either from a pilot project in the case of a new program, or from experience with a program being evaluated), these could be used rather than adopting the approach outlined above.

Avoidance of incorrect payments

This may occur, for example, where a data-matching program identifies that someone currently receiving a Government payment is not entitled to it, or is not entitled to payment at the current rate, and this leads to termination or reduction of the payment.

Especially in the case of continuing payments (for example, welfare benefits) it will often not be possible to conclusively determine how much would have been incorrectly paid had the payment not been terminated or reduced.

If a general assumption is made (for example, that the incorrect amount would have continued to be paid for a standard period), the statement should clearly specify the assumption, and the reasons for its adoption. For example, an agency might adopt assume that an incorrect payment would have continued to be paid for half the average period for which payments of that type are made, based on past experience regarding the average time required to identify incorrect payments.

In making such an assumption, agencies should account for the possibility that other review methods (if relevant or applicable) could have identified the incorrect payment had the data-matching program not done so.

Increasing the revenue collected by an agency

If a program identifies cases where additional revenue is owed to an agency, the estimate of the benefit derived should either:

- allow for the possibility that all revenue owed will not be collected, and the cost of collection (unless the cost of collection is included as a cost of the program), or
- be based on the actual amounts collected as a result of the data-matching program and the costs of collection.

Indirect financial benefits

Administrative savings

Savings of this sort are likely to be most important where data-matching allows an activity that would have to be carried out in any case, to be performed more efficiently.

One approach to estimating these savings would be to estimate the comparative cost of carrying out the activity with and without data-matching (see the section above titled 'Methods of presenting costs/benefit information').

Voluntary compliance

Agencies may think that public knowledge that a data-matching program is operating leads to increased compliance with the law, thus reducing regulatory costs. This kind of benefit is obviously difficult to quantify. If agencies believe that benefits of this sort are likely to be achieved, they should include them in the statement of benefits, along with the reasons for holding this view and any information that indicates the likely magnitude of benefits from this source. This would require estimates of current error rates or fraud rates together with an assessment of the program's anticipated impact on them.

Non-financial benefits

Many data-matching programs have benefits that cannot readily be expressed in financial terms. For example, data-matching is used to locate illegal immigrants, to detect criminal offences and to build up intelligence holdings of law enforcement agencies.

Some types of benefits probably cannot be quantified at all, but should still be described. For example, if a benefit of data-matching is improved service to clients or improved data quality, the statement of benefits could describe the effect of the data-matching program in these regards.

Where possible, it is helpful to quantify non-financial benefits. For example, if a program is aimed at locating illegal immigrants, it is useful to state how many illegal immigrants have been, or are expected to be, located by means of the program. This helps to illustrate the reasons why a data-matching program is considered worthwhile, and provides a basis for comparing actual performance against initial estimates.

If a data-matching program does not have a readily quantifiable outcome of this sort, other measures of performance can be found. For example, if the function of a

program is to add significant items of information to an intelligence database, it may be relevant to estimate how many items of information the program will identify. If matches contribute to an outcome but are not the sole factor, it may be useful to indicate in how many instances the output from the data-matching program contributes to a result being achieved.

Further ideas on how to present cost and benefit information for data-matching programs that have quantifiable non-financial benefits are presented above under the heading 'Methods of presenting costs/benefit information'.