The use of data matching in Commonwealth administration - Guidelines

February 1998

Contents

1 Introduction

1 These guidelines have been developed in consultation with Commonwealth agencies to assist in ensuring that data-matching programs are designed and conducted in accordance with sound privacy practices.

2 They are issued by the Privacy Commissioner under s.27(1)(e) of the Privacy Act, which provides that the functions of the Commissioner include:

to prepare and to publish in such manner as the Commissioner considers appropriate, Guidelines for the avoidance of acts or practices of an agency that may or might be interferences with the privacy of individuals or which may otherwise have any adverse effects on the privacy of individuals.

3 The Privacy Commissioner issued data-matching guidelines for discussion in 1990 and guidelines for adoption by agencies in 1992. The 1992 guidelines were revised and re-issued in November 1995. This document contains a revision of the 1995 guidelines, addressing a number of practical problems raised by agencies in light of experience. Further information on adoption of and compliance with the guidelines is available in the Privacy Commissioner's Annual Reports.

Operational advantages and privacy risks of data-matching

4 Data-matching is a powerful administrative and law enforcement tool. It allows information from a variety of sources to be brought together, compiled and applied to a range of public policy purposes at vastly lower cost than manual methods.

(i) In benefit paying programs it can be used to check the information clients provide and to identify clients who are receiving an incorrect level of benefits.

(ii) In revenue collection it can allow identification of undeclared income and help in the efficient allocation of audit resources.

(iii) In criminal investigation it can provide vital intelligence and show up otherwise invisible links between persons and investigations.

(iv) Data-matching can play an important role in reducing fraud on the Commonwealth. For many types of fraud it is one of the few instruments available for determining that an offence has been committed at all.

5 All of these are important public policy purposes and data-matching can make a valuable contribution to achieving them.

6 At the same time, it can pose risks to the privacy of the people whose data is being matched.

(i) It may involve the use of data for purposes other than those for which it has been supplied or obtained and those purposes may be outside the reasonable expectations of the people the information is about. A basic privacy principle is that personal information should be used only for the purpose for which it was obtained. Departures from this principle need to be justified on strong public interest grounds.

(ii) Data-matching can involve the automatic examination of the personal information of many thousands of people about whom there are no known grounds for suspicion and in relation to whom no action is warranted. This may be done without the knowledge of the people whose information is being scrutinised.

(iii) Data-matching relies on agencies gaining access to large amounts of information, some of which may be personal information, from other sources. Agencies may be inclined to keep unmatched information for possible future use even though it has no immediate application.

(iv) It is far from perfectly reliable. A data-matching program may fail to distinguish between individuals with similar personal details; input data may be faulty; errors may be made in programming; or difficulties may be caused if similar fields in different databases are not precisely comparable.

7 The guidelines seek to address these issues by ensuring that wherever possible:

(i) the estimated costs and benefits of data-matching programs are taken into account before commencement;

(ii) information about the data-matching programs in which Commonwealth agencies participate is made publicly available;

(iii) people whose data is used for data-matching programs are informed about those uses;

(iv) the output of data-matching programs is not accorded undue weight and is checked before action is taken which could be detrimental to the individual concerned;

(v) data obtained for use in a data-matching program is not retained once it has served the specific purposes of that program; and

(vi) the effectiveness of data-matching programs is carefully examined.

Status of the guidelines

8 These guidelines are not legally binding. The Privacy Commissioner intends to seek agreement from Commonwealth agencies to adopt the guidelines in relation to data-matching programs in which they participate. An agency which has given such an agreement would not be acting unlawfully if it did not abide by the guidelines.

9 An agency would not fail to comply with the guidelines by virtue of taking any action which is required or specifically authorised by law.

10 While the Privacy Commissioner may take these guidelines into consideration in assessing compliance with the Information Privacy Principles (IPPs) in section 14 of the Privacy Act, these guidelines aim to encourage a higher standard of regard for people's privacy rights in relation to data-matching than is required by bare compliance with the IPPs and an agency would not necessarily breach the IPPs if it did not adhere to these guidelines.

11 The Privacy Act provides the Privacy Commissioner with the power to audit agencies, to investigate complaints and to undertake investigations on his or her own account. The guidelines recognise both the management responsibilities and accountability of agencies and the monitoring, audit, enforcement and advisory powers of the Privacy Commissioner. Adoption of these guidelines by agencies reflects both good management practices and agencies' commitment to the protection of individuals' privacy rights.

12 Agencies are responsible for applying the guidelines to their data-matching activities and for determining whether it is appropriate to apply particular requirements to particular data-matching programs. However the guidelines include a provision that agencies should formally advise the Privacy Commissioner when they propose to depart from the normal safeguards of the guidelines.

13 The date of effect of these guidelines is 1 April 1998.

2 Definitions

14 The definitions in the Privacy Act apply to these guidelines, as do the following.

Administrative action

Action taken as result of a data-matching cycle or program that materially affects any individual or class of individuals, including:

(i) taking any action directly detrimental to the individual, such as reducing a benefit or imposing a penalty;

(ii) initiating an investigation which might lead to detrimental action; and

(iii) disclosing information to any third party where the disclosure might cause harm or embarrassment to the individual.

Database

A collection of information stored in a computer and organised in categories to facilitate retrieval.

Data-matching

The large scale comparison of records or files of personal information, collected or held for different purposes, with a view to identifying matters of interest.

Data-matching cycle

The conduct of all steps and processes allowed by a data-matching program within a specified timeframe.

Data-matching program

A set of one or more data matching cycles which are consistent as to matters such as the data which is accessed and the matching criteria applied, and which are undertaken to assist one or more organisations in addressing a definable objective.

Match

A result obtained in a data-matching program in relation to which further administrative action may be taken by the matching agency or source agency or organisation; this could either be a positive match of like details or a meaningful non match or discrepancy.

Matching agency

In relation to a data-matching program, the agency on whose computer facilities the matching is conducted.

Primary user agency

In relation to a data-matching program, the agency that makes the most substantial use of the program's results. Usually the primary user agency will also be the matching agency, but there will be some programs where the matching is conducted on the computer facilities of an agency that either does not use, or uses only to a minor extent, the results of the program. Where there is more than one agency using the results of a program, user agencies should agree which is the primary user.

Source agency or organisation

Any agency governed by the Privacy Act or any other body (including a State or local government agency or a private sector or voluntary organisation) which discloses information contained in a system of records to a matching agency for use in a data-matching program.

Source data

The computer record from which the extract provided for data-matching was taken; paper file; or original application.

User agency

Any agency that uses the results of a data-matching program.

3 Coverage and application

Coverage of the guidelines

15 These guidelines apply to a data-matching program if none of the exclusions in paragraph 16 applies and all of the following criteria are met:

(i) the program involves the computerised comparison of two or more databases or extracts of databases, and at least two of the databases or extracts contain information about more than 5,000 individuals; and

(ii) the information in the different databases was collected for different purposes; and

(iii) the purpose of the program is:

(A) to select individuals for possible further administrative action; or

(B) to add information from one database to another for purposes which include taking administrative action in relation to the individuals concerned; or

(C) 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

(D) to combine permanently the databases being matched.

16 These guidelines do not apply to a program if:

(i) its objective is to verify information provided by an individual about the individual's circumstances, status or relationship (as recorded by an agency) with another agency or organisation, and the information will not be used materially in the making of a decision to take administrative action; or

(ii) its objective is to verify information provided by an individual about the individual's circumstances, status or relationship (as recorded by an agency) with another agency or organisation, and the information may be used materially in the making of a decision to take administrative action, and the program is conducted within three months of the individual providing the information; or

(iii) its objective is to co-locate records or data items previously held in separate locations and the co-location does not result in any change to the purposes for which the records or data items are used or disclosed.

17 If an agency runs a number of very similar programs (for example, programs which have one source database and the algorithm in common, but vary as to the other source database), they should be treated as a single program for the purpose of assessing whether they are subject to the guidelines. In particular, the other source databases should be regarded as a single database when deciding whether the databases used in the program contain records about more than 5,000 individuals.

18 If an agency conducts a number of similar programs that have the same objective and allow the drawing of similar inferences about the individuals identified, the agency may treat these as a single program for the purpose of meeting the requirements of these guidelines.

19 Since different agencies generally hold information for different purposes, most data-matching programs in which a number of different agencies participate will

satisfy the condition set out in paragraph 15(ii). However, where two or more agencies have joint responsibility for managing a government program it is possible that databases of information collected for the same purpose would be held by different agencies. As the information was not collected for different purposes, matching these databases would not be covered by the guidelines.

Programs not covered by the guidelines

Requirement to report to the Privacy Commissioner

20 On an annual basis, each agency should prepare and provide to the Privacy Commissioner a report on data-matching programs for which it is the matching agency, which involve databases that each have more than 1,000 records, and which would be covered by the guidelines if not for paragraph 15(i).

21 On request by the Privacy Commissioner, each agency should prepare and provide to the Privacy Commissioner a report on data-matching programs for which it is the matching agency, which involve databases that each have more than 1,000 records, and which would be covered by the guidelines if not for paragraph 16(ii).

22 Where feasible, a report prepared in accordance with paragraph 20 or 21 should list each program.

23 Where many such data-matching programs are undertaken on an ad hoc basis and it is not feasible to list each one, a description of the data-matching activity should be prepared and provided to the Privacy Commissioner, covering what sort of purposes the data-matching serves (including what is done with the results), how much of it is being conducted, what data is involved and what categories of staff within the agency carry it out.

Privacy safeguards

24 If in the course of participating in a data-matching program not covered by these guidelines an agency becomes aware of information that it considers necessitates investigative or administrative action to be taken against an individual, it should accord that individual the safeguards set out in paragraphs 63 to 68. This applies to each agency using the results of the program.

Data-matching with non-Commonwealth organisations

25 A Commonwealth agency which is involved with a non-Commonwealth organisation in a data-matching program, whether as the matching agency or a source

agency, should seek by negotiation with the other party to ensure that the requirements of these guidelines are adopted in relation to the program.

Programs inconsistent with the requirements of the guidelines

26 Where the head of an agency considers that it would be appropriate in the public interest to conduct a data-matching program in a way inconsistent with one or more of the requirements of the guidelines (as set out in Part 4), he or she should:

(i) formally advise the Privacy Commissioner of the details of the proposed program;

(ii) in that advice, specify those requirements of the guidelines with which the proposed program would be inconsistent; and

(iii) in that advice, explain the public interest grounds for the decision.

27 The Privacy Commissioner's normal practice will be to make advice under paragraph 26 publicly available. The Privacy Commissioner will keep such advice confidential if, in its advice under paragraph 26, the agency requests the Commissioner to do so and explains its reasons for making that request.

28 In providing advice under paragraph 26(iii), the head of the agency should address such matters as:

(i) the effect that not abiding by the specified requirements of the guidelines would have on individual privacy;

(ii) the seriousness of the administrative or enforcement action that may flow from a match;

(iii) the effect that not abiding by the specified requirements of the guidelines would have on the fairness of the program - including its effect on people's ability to find out the basis of decisions that affect them and their ability to dispute those decisions;

(iv) the effect that not abiding by the specified requirements of the guidelines would have on the transparency and accountability of government operations;

(v) the effect that not abiding by the specified requirements of the guidelines would have on compliance of the proposed program with the Information Privacy Principles in the Privacy Act;

(vi) the effect that abiding by all the requirements of the guidelines would have on the effectiveness of the proposed program;

(vii) whether abiding by all the requirements of the guidelines could jeopardise, or endanger the life or physical safety of, information providers or could compromise the source of information provided in confidence;

(viii) the effect that abiding by all the requirements of the guidelines would have on public revenue - including tax revenue, personal benefit payments, debts to the Commonwealth and fraud against the Commonwealth;

(ix) whether abiding by all the requirements of the guidelines would involve the release of a document that would be an exempt document under the Freedom of Information Act 1982; and

(x) the legal authority for conducting the proposed program in a way inconsistent with the specified requirements of the guidelines.

29 The Privacy Commissioner may respond to the agency's advice, setting out:

(i) the Commissioner's view as to whether it would be appropriate, from a privacy protection perspective, for the specified requirements of the guidelines not to be followed; and

(ii) the Commissioner's reasons for taking that view.

30 If the Commissioner takes the view that it would be inappropriate, from a privacy protection perspective, for the specified requirements of the guidelines not to be followed, the Privacy Commissioner may suggest changes which would, if made, satisfy him or her that the proposed program offers an adequate standard of privacy protection.

31 It should be noted that:

- paragraph 27(1)(k) of the Privacy Act provides that it is a function of the Privacy Commissioner:

to examine (with or without a request from a 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 the adverse effects of such proposed enactment on the privacy of individuals are minimised;

- subsection 32(1) of the Privacy Act provides:

Where the Commissioner, in the performance of the function referred to in paragraph 27(1) ... (k), has monitored an activity or conducted an audit, the Commissioner may report to the Minister [the Attorney-General] about that activity or audit, and shall do so if so directed by the Minister; and

- subsections 31(2) and (3) provide:

(2) If the Commissioner thinks that the proposed enactment [a proposed enactment examined by the Commissioner under paragraph 27(1)(b) of the Privacy Act] would require or authorise acts or practices of an agency that would be interferences with the privacy of individuals, the Commissioner shall

(a) report to the Minister about the proposed enactment; and

(b) include in the report any recommendations he or she wishes to make for amendment of the proposed enactment to ensure that it would not require or authorise such acts or practices.

(3) Otherwise, the Commissioner may report to the Minister about the proposed enactment, and shall do so if directed by the Minister.

4 Requirements of the guidelines

Proceeding with a program

32 Before a Commonwealth agency decides to participate in a new data-matching program, or to recommend that such a program should commence, it should take account of the following matters:

(i) the estimated costs and benefits of the program (see also Appendix D); and

(ii) any alternative measures to data-matching which could achieve the same results as the proposed matching program.

Public notice

Requirement to give public notice

33 Before participating in a data-matching program covered by these guidelines, each agency involved should ensure that a reasonable form of public notice of the proposed program is given.

34 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 Privacy Commissioner, public notice of the program is given in accordance with paragraphs 36 to 40.

35 Where a number of similar programs have been treated as a single program under paragraph 18, the Gazette notice (see paragraph 36) and other publicity material should deal with the matters set out in paragraph 36 and should also describe the range of data sources used, how many different programs are involved and what classes of agency staff are responsible for conducting them.

Forms of public notice

36 The primary user agency in relation to a proposed data-matching program should give public notice of the program in the Gazette. The notice should contain:

(i) brief details of the objectives of the program;

(ii) the agencies and categories of data involved;

(iii) the classes of individuals about whom personal information is to be matched, and the approximate number of individuals affected.

37 The Gazette notice should be published before the program commences.

38 The primary user agency should forward copies of the Gazette notice to individuals or organisations nominated by the Privacy Commissioner.

39 Each agency participating in a program in any capacity should take steps to publicise the program as widely as is practical. This should include both steps to inform the general public about the program (for example the Gazette notice and newspaper advertisements) and steps specifically aimed at informing people whose information is likely to be used in the program, for example:

(i) including information about the matching program in material given to people when they provide information which is likely to be used in the program;

(ii) informing relevant clients about the program directly; and

(iii) putting notices in relevant special-purpose publications or newsletters.

40 The Gazette notice and other publicity material should advise how the general public can obtain copies of the program protocol.

Entry in the Personal Information Digest

41 A matching agency should include in its PID entry a description of any records of personal information it holds in connection with a data-matching program. A source agency should include in the relevant part of its PID entry a note that the records are disclosed to the matching agency in connection with data-matching.

Program protocol

Requirement to prepare, make publicly available and observe

42 Where a data-matching program has not yet commenced, the primary user agency should prepare a program protocol. A copy should be forwarded to the Privacy Commissioner before the program is due to commence. The protocol should be made publicly available at the commencement of the program.

43 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 Privacy Commissioner, a protocol is prepared, forwarded to the Privacy Commissioner and made publicly available.

44 Where a number of similar programs have been treated as a single program under paragraph 18, the protocol prepared to cover those programs should deal with the matters set out in paragraph 46 and should also set out each data source used, how many different programs are involved and what classes of agency staff are responsible for conducting them.

45 Each agency participating, in any capacity, in a data-matching program should ensure that its participation in the program is in accordance with a program protocol.

Contents

46 The program protocol should consist of a description of the program and an explanation of the reasons for undertaking it. The description of the program should include:

(i) an overview of the program;

(ii) the objectives of the program;

(iii) the matching and source agencies, and any agencies which will use the results of the program;

(iv) a description of the data to be provided and the methods used to ensure it is of sufficient quality for use in the program;

(v) a brief description of the matching process, the output produced and the destination of the results of the program;

(vi) what action, administrative or otherwise, may be taken as a result of the program;

(vii) time limits applying to the conduct of the program

(viii) what form of notice has been given, or is intended to be given, to individuals whose privacy is affected by the program; and

And, the explanation of the reasons for undertaking the program should include:

(ix) the program's relationship to the agencies' lawful functions and activities;

(x) the legal authority for the uses and disclosures of personal information involved in the program;

(xi) alternative measures to data-matching which were considered, and the reasons why they were rejected;

(xii) information about any pilot testing of the program; and

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

47 A suggested format for presentation of the protocol is at Appendix A.

Technical standards report

Requirement to prepare and observe

48 Detailed technical standards should be established by the matching agency, where practicable in consultation with source agencies or organisations, to govern the conduct of the program, and be set out in a Technical Standards Report.

49 Each agency participating, in any capacity, in a data-matching program should ensure that its participation in the program is in accordance with the Technical Standards Report.

50 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.

51 The matching agency should lodge a copy of the Technical Standards Report with the Privacy Commissioner on request.

52 These guidelines do not require the Technical Standards Report to be made publicly available.

53 For new programs, the 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 initial experience of the operation of the program.

54 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 Privacy Commissioner.

55 The Technical Standards Report should be amended whenever the specifications for the program change.

56 The Technical Standards report will assist in the proper audit of an agency's compliance with the Information Privacy Principles in that it presents information in a manner which is capable of independent scrutiny.

57 As well as meaning that an agency has not adhered to these guidelines, failure to implement practices relating to data quality of the kind indicated in paragraph 58 may also lead the Privacy Commissioner to find that an interference with privacy is occurring or has occurred because of non-compliance with IPPs 4, 7, 8 or 9.

Contents

58 The report should deal with the following matters:

(i) description of data supplied by source agencies or organisations, referring in particular to: key terms and their definition, relevance, timeliness and completeness;

(ii) the specification for each matching algorithm or project should be outlined and include such things as:

- the risk to be addressed;
- data items used in the match, in particular the use of any identification numbers;
- the rules for recognising the match; and
- the destination of the results;

(iii) controls being employed to ensure the continued integrity of the program and its data; and

(iv) security features included within the program to control and minimise access to personal information.

59 Further guidance in relation to the contents of the Technical Standards Report is given in Appendix C.

Monitoring by the Privacy Commissioner

60 Agencies should enable the Privacy Commissioner to undertake inspections of their data-matching activities and their procedures. The monitoring process carried out by the Privacy Commissioner will include assessing the outcomes of the program from a privacy perspective, assessing whether the program is being conducted in accordance with the procedures set out in the program protocol, reviewing the effectiveness of the controls and procedures set out in the Technical Standards Report and considering any complaints and difficulties that have arisen in connection with the program.

61 Agencies should report to the Privacy Commissioner, if requested, on any relevant matter, including:

(i) actual costs and benefits flowing from the program;

(ii) any non financial but quantifiable factors that are considered relevant;

(iii) difficulties in the operation of the program and how these have been overcome;

(iv) the extent to which internal audits or other forms of assessment have been undertaken by agencies or organisations and their outcome; and

(v) 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.

62 The Privacy Commissioner proposes to include in the Commissioner's annual report general information about:

(i) the number, extent and nature of data-matching programs;

(ii) the extent of public notification of programs and of consultation;

(iii) the extent of confidential notification of programs to the Privacy Commissioner;

(iv) the nature of the public interest reasons advanced for not engaging in public notification of programs; and

(v) the operational experience and effectiveness of programs.

Handling data-matching information

Informing individuals

63 An agency should only take administrative action as a result of a match, including conducting checking with third parties, after giving the individual at least 14 days to consider the information and comment on its contents. If there is a dispute as to the accuracy of the data or the proposed administrative action but the agency considers that further action is justified, it should inform the individual of the rights conferred by the Privacy Act to lodge a complaint with the Privacy Commissioner.

64 If an agency has advised the Privacy Commissioner (in line with paragraph 26) that it plans to take administrative action as a result of a match without checking the results with the individual concerned, it should take whatever other steps are reasonable to check the information (for example, checking by reference to the source data) before taking action.

65 It should be noted that the decision as to further action should ideally involve individual consideration of each case. However, where a program is generating large numbers of matches, it may be acceptable to apply some rule based selection criteria, which should be described in the Technical Standards Report. The objective of the requirement is to achieve further filtering or selection of cases independently of the matching program itself.

66 Unless otherwise provided by law, an agency should take no administrative action that interferes with the individual's opportunity to exercise any rights of appeal or review.

67 Paragraphs 63 to 66 apply to each agency using the results of the matching program.

68 It should be noted that, as well as meaning that an agency has not adhered to these guidelines, failure to implement practices relating to data quality of the kind indicated in paragraphs 63 and 64 may also lead the Privacy Commissioner to find that an interference with privacy is occurring or has occurred because of non-compliance with Information Privacy Principles 7, 8 or 9 in section 14 of the Privacy Act.

Record controls

69 Personal information obtained for use in a matching program which does not lead to a match should be destroyed by the matching agency as soon as practicable

after the relevant matching process has been completed. In any case the destruction of the information should not occur any later than 90 days after the matching process.

70 In cases where a match occurs a decision as to further action in relation to each individual should be taken by an agency within 90 days of the matching process. If, during that time or any later time, a decision not to take further action is made, the information should be destroyed, where practicable within 14 days but at least within 90 days. This applies to each agency using the results of the matching program.

71 The requirements in paragraphs 69 and 70 are consistent with the disposal arrangements for records relating to data-matching exercises set out in Australian Archives General Disposal Authority 24, issued on 13 January 1997.

Limitation on new registers

72 An agency involved, in any capacity, in a data-matching program should not create any new separate permanent register (or database) from information about individuals whose records have been matched as part of the program.

73 Paragraph 72 does not prevent:

(i) the maintenance of a register of individuals in respect of whom further investigations are warranted under the terms of the program protocol and following a decision under paragraph 70; or

(ii) the maintenance of a special register solely for the purpose of excluding individuals from being selected for investigation in successive cycles of the same matching program.

74 Paragraph 72 does not affect the addition of updating information to an existing record, provided no other administrative action is intended.

75 Paragraph 72 is aimed at preventing agencies compiling dossiers or otherwise collating information obtained only through a data-matching program. It does not prevent an agency from maintaining such records or databases as are reasonably necessary to the fulfilment of the objectives of the program.

Evaluation

76 Not later than three years after the commencement of operation of a data-matching program, and at least every three years after that while the program continues, the primary user agency should undertake an evaluation of the program in accordance with its original objectives. The outcome of this should be documented by

the matching agency, and should include a statement of the costs and benefits of the program (see Appendix D for a description of what such a statement should contain). A report of the evaluation should be given to the Privacy Commissioner. The report should be made publicly available, unless it refers to a program of which public notice was not originally given.

Changes to data-matching programs

77 When a matching agency wishes to make minor changes or amendments to a data-matching program these should be noted on the program protocol document, advised to the Privacy Commissioner and considered in the program evaluation.

Appendix A: Content of data-matching program protocols

The program protocol should provide two main things:

- a description of the program; and
- an explanation of the reasons for deciding to conduct the program.

The function of the program protocol is to inform the public about the existence and nature of the data-matching program. It should therefore be written as plainly as is consistent with giving an accurate picture of how the program works.

A suggested format is set out below.

Description of the program

The description of the program needs to cover the following matters. Possible section titles are in parentheses.

(i) An overview of the program. (Overview)

This should be a short, simply expressed statement of what the program does and why. A half a page should be enough.

(ii) The objectives of the program. (Objectives)

This should be a basic statement of what the program is trying to achieve. The rest of the protocol will flesh out exactly how the program tries to achieve its objectives so this statement need not be a long one.

(iii) The matching and source agencies, and any agencies which will use the results of the program. (Agencies involved)

This should include:

- what agency is conducting the matching program;
- where the matching agency is not the same as the primary user agency, the protocol should make clear which agency does what;
- what agencies or other sources are providing data which will be used in the 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 program.

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

For each data or information source involved in the program, give a brief description of:

- the 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.

(v) A brief description of the matching process, the output produced and the destination of the results of the program. (The matching process)

This should describe which fields are matched (eg Medicare number, 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 fine details of the matching will be set out in the Technical Standards Report, so the description here may be kept in relatively broad terms.

(vi) What action, administrative or otherwise, may be taken as a result of the 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 customer contact letters should be attached to the program protocol as an appendix.

(vii) Time limits applying to the conduct of the program. (Time limits)

This should cover:

- how long data obtained for use in the 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 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 program, including agencies' internal review mechanisms as well as external mechanisms such as legislative sunset clauses.

(viii) What form of notice has been given, or is intended to be given, to individuals whose privacy is affected by the program. (Public notice of the program)

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

Reasons for deciding to conduct the program

The reasons for deciding to conduct the program should cover the following issues.

(ix) The program's relationship to the agencies' lawful functions and activities. (Relationship to lawful functions)

This should be a statement of the reasons for considering that the program will improve public administration or otherwise serve the public interest.

(x) The legal authority for the uses and disclosure of personal information involved in the program. (Legal authority)

The use and disclosure of personal information should be justified in terms of the Privacy Act and any other relevant legislation.

To be lawful, any disclosure of personal information by an agency must fall within one of the five exceptions in Information Privacy Principle 11 in the Privacy Act.

Similarly, any use of personal information by an agency for a purpose other than the purpose for which the agency collected it must fall within one of the five exceptions to Information Privacy Principle 10 in the Privacy Act.

The protocol must specify which exceptions in Principles 10 and 11 apply, and why.

If there are other provisions in legislation relevant to the use and disclosure of information involved in the program, for example secrecy provisions, the protocol should explain how the uses and disclosures are authorised in terms of those provisions.

(xi) Alternative measures to data-matching which 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.

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

Information about a pilot project which is likely to be relevant includes:

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

If the protocol relates to a new program (rather than one already operating) and no pilot project has been conducted or is planned, the protocol should indicate why no pilot is considered necessary.

(xiii) A statement of the costs and benefits of the program. (Costs and benefits)

See Appendix D for what should be included here.

Appendix B: Sample program protocol

What follows is an example of a program protocol addressing the requirements of the voluntary guidelines. It is based on an actual protocol produced by the Department of Social Security but, because the purpose it serves here is to illustrate the main features needed in a protocol (rather than to present the full range of detail that might be included on a particular program), it has been somewhat abbreviated for accessibility and ease of reading. Attachments referred to in the sample protocol have been omitted.

Program protocol: Matching Information from the Australian Taxation Office's Employment Declaration Form with Information on Customers and Customers' Spouses from the Department of Social Security

Description of the program

1 Overview

The Privacy Commissioner's Guidelines on Data-matching in Commonwealth Administration specify that a program protocol be prepared by agencies conducting more significant data-matching programs. The Department of Social Security (DSS) has agreed to comply voluntarily with the guidelines, which are not legally binding.

Pay As You Earn (PAYE) employees must fill in an Employment Declaration Form (EDF) if they are to avoid being taxed at the highest marginal rate. If EDF data indicates that a DSS customer (or spouse) is employed, but the earnings from that employment have not been disclosed to DSS, it is likely that the person has been receiving a level of DSS benefits to which he or she is not entitled.

In the 1991 Budget, the Government introduced arrangements under which the Australian Taxation Office (ATO) provides all EDF information to DSS weekly on magnetic tape for matching with DSS's customer income information to allow the identification of incorrect DSS payments. In the 1993 Budget, the Government extended the matching of EDF data to DSS information on the income of the spouses of Newstart system customers (Jobsearch Allowance, Newstart Allowance, Sickness Allowance and Special Benefit recipients).

DSS's Security, Fraud and Control Division receives data from the ATO, matches it with DSS data, ensures the security of the data during processing and destroys unmatched data within 90 days of receiving the data from the ATO.

Cases identified are forwarded to DSS regional offices for review.

2 Objectives

The objectives of the programs are:

- to achieve savings in DSS outlays by identifying customers or spouses of customers who are not disclosing income from employment;
- to recover outstanding payments by locating people on the DSS's debtor file who are currently employed (and therefore may be able to repay their debts); and
- by publicising the data-matching, to promote voluntary compliance with income disclosure requirements.

From a broader perspective, achievement of these goals will further social justice objectives by providing additional scope for the direction of funds to those in most need.

3 Agencies involved

DSS is the matching agency: the data-matching program runs on its computer facilities.

DSS contributes data relating to current or suspended recipients of Newstart Allowance, Job Search Allowance, Sickness Allowance, Special Benefit, Sole Parent Pension and Widow's Pension and to spouses of current or suspended recipients of Newstart Allowance, Job Search Allowance, Sickness Allowance and Special Benefit.

The ATO data relates to EDF information, including: name; date of birth; home address; residency status; full time, part time or casual employment status; rebates claimed; employer's business name; employers postal address; and the nature of employers business. The ATO data does not include Tax File Numbers. Around 220,000 records are provided by the ATO each fortnight.

Health Insurance Commission data is used for data quality purposes (see section 4 below).

DSS is the only agency that uses the output of the program (see section 6).

4 Data issues

Data quality. DSS requires that information about changes in customers' circumstances (eg, change of address, income received, marital status, family membership) are recorded on DSS master files. Customers are required to advise changes of circumstances within 14 days of an event occurring. Details can be checked when a customer's entitlements are reviewed. Data items from the ATO are checked before matching to ensure that the data items contain logically feasible information.

Data integrity. Some EDF information from the ATO has only the year of birth. In these cases, surname, forename, middle initial, sex and year of birth are matched

against Health Insurance Commission data. If there is exactly one match, the full date of birth from the HIC record is deemed to refer to the same individual as the EDF and is added to the EDF data.

Security. DSS staff are subject to existing security controls and the confidentiality provisions of the Social Security Act. Access to the computing centre is strictly controlled and entry properly authorised. The ADP security system provides protection and control of dataset access and system entry and program integrity. Features include logon identification codes, passwords and security groupings to ensure that access to information is on a needs only basis. The ATO computer cartridges are scheduled and released by ATO; they are logged in and out of the ATO and DSS. Appropriate data destruction practices are in place (see section 7). Existing security arrangements in DSS automatically log user access to data files. Each program in the EDF matching cycle produces run statistics on number of records read, number of records matched, number of records unmatched and number of records written. Each program also provides audit trails which can be used to verify the processing of the data.

5 The matching process

The matching identifies customers who appear not to have disclosed income earned as an employee. People who appear on both files are considered to be ‘matches’.

6 Action resulting from the program

At the completion of the matching, selected customer and employer information is loaded into the Department’s National Selective Review System for release to the appropriate DSS regional office for review. The current status of the customer is checked on the DSS payment system and the customer is contacted by phone, letter or visit to advise that information has been received indicating that an EDF was signed by the customer (or his or her spouse) and that the income details from this employment are required. The customer, therefore, has an opportunity to respond to the information before any action is taken to confirm the information with a third party (usually the employer). DSS will not interrupt payments until the information has been checked with the customer. Income details provided by the customer may be checked with the employer.

The customer’s entitlement is reviewed. Depending on the result, further action may be taken, including downward variation of payment, cancellation of payment, raising an overpayment and, in more serious cases, prosecution.

7 Time limits applying to the conduct of the program

Data is destroyed in accordance with the Privacy Commissioner’s guidelines Data-matching in Commonwealth Administration: ATO data deemed unsuitable for matching is destroyed within 14 days; ATO data used in the matching run but not matched is destroyed within 90 days; and all remaining ATO data is destroyed within 12 months.

The matching is done fortnightly.

8 Public notice of the program

The EDF form includes a statement that 'Some information provided on this form may be given to certain government agencies as described in taxation law.'[1]

DSS provides customers with general information about its data-matching on Regional Admission Procedures cards. Leaflets on data-matching were included in a national mail out to Jobsearch and Newstart Allowance customers in June 1994. Articles on data-matching have been published in Aged Pension News, Impact, and media releases during 1993 and 1994.

Reasons for deciding to conduct the program

9 Relationship with the agencies' lawful functions and activities

The program is clearly related to DSS's lawful function of limiting payments to those eligible under relevant legislation. The Social Security Act provides that customers are required to disclose their other income (and, where relevant, their partner's income). These requirements are given to customers in a written advice authorised under different sections of the Act for different payment types; for instance, the authority in relation to Job Search Allowance appears in section 574.

10 Legal authority

Paragraph 16(4)(e) of the Income Tax Assessment Act allows the ATO to pass information to DSS for the purpose of administering any Commonwealth law relating to pensions, allowances or benefits.

Under subsection 130(7) of the Health Insurance Act, the Minister for Health may issue a certificate authorising the release of Health Insurance Commission information for the purpose of the administration of an Act administered by the Minister for Social Security. Such a certificate has been issued by the Minister in relation to this program.

These provisions mean that the disclosures involved are authorised by law and consequently that they are lawful under Information Privacy Principle 11 in s.14 of the Privacy Act, which provides that an agency in control of personal information shall not disclose it unless one of a number of exceptions (including that the disclosure is required or authorised by law) applies.

11 Alternative methods

Before the changes announced in the 1991 Budget, DSS received the EDF data on paper (except for Queensland data, which came on tape). A selection of the forms was keyed in by DSS and the forms returned to the ATO. This method was not timely because of the delay between when the employee signed the form and when the actual matching took place. Under the 1991 initiative, processing of the electronic data has

[1] [Note from the Privacy Commissioner: the Commissioner does *not* regard this statement as meeting the requirements of Information Privacy Principle 2 in the Privacy Act.]

proven to be cheaper and more timely than the system it replaced as well as increasing the savings from the program.

Comparisons with other methods show that alternatives like Mobile Review Team activity, public denunciation and industry reviews are more time consuming and labour intensive than data-matching.

12 Pilot programs

As part of the May 1987 Economic Statement, DSS began matching a proportion of EDF information against DSS customer records on a trial basis. Until this program began in 1991, EDF information from Queensland and Western Australia was provided by the ATO on magnetic tape while EDF data from other States were sent to DSS in paper form and a sample of them matched with DSS records. Since this program is an extension of that partial matching system, which had proved effective within its considerable technical constraints, no formal pilot testing was undertaken.

As set out in section 6 above, the customer EDF matching started in December 1991 and was evaluated in a Post Implementation Review (PIR) report in December 1992. The spouse matching started in October 1993 and the program was evaluated in a PIR report in September 1994.

13 Costs and benefits

The customer EDF matching started in December 1991 and was evaluated in a Post Implementation Review (PIR) report in December 1992. The spouse matching started in October 1993 and the program was evaluated in a PIR report in September 1994.

The following table sets out experience to date with the 1991 initiatives (the matching with DSS customer information).

	1991-92	**1992-93**	**1993-94**	**1994-95**
Selections (number)	89,896	149,795	163,366	266,378
Reviews completed (number)	59,614	114,649	133,013	218,621
Downward variations, incl cancellations (no.)	3,241	7,751	9,297	16,221
Reduction in fortnightly outlays ($m)	0.767	1.565	1.991	3.141
Overpayments detected (number)	12,219	29,500	42,932	62,412
Overpayments detected ($m)	7.7	21.9	32.2	50.4
Overpayments expected to be recovered ($m)[2]	5.4	15.4	22.6	35.3
Actual savings ($m)[3]	11.2	49.8	68.8	117.1

Savings carry over to next year ($m)[4]	14.1	20.4	25.9	40.8
Cost of reviews ($m)	3.8	7.6	10.3	16.1
Cost:benefit ratio	2.93	6.58	6.22	7.26

The following table sets out experience to date with the 1993 initiatives (the matching with DSS customer spouse information).

	1993-94
Selections (number)	14,141
Reviews completed (number)	10,426
Downward variations, incl cancellations (number)	854
Reduction in fortnightly outlays ($m)	0.299
Overpayments detected (number)	3,883
Overpayments detected ($m)	5.0
Overpayments expected to be recovered ($m)	3.5
Actual savings ($m)[5]	6.4
Savings carry over to next financial year ($m)[6]	4.9
Cost of reviews ($m)	1.0
Cost:benefit ratio	6.39

People claiming or receiving income support payments are more likely to comply voluntarily with the law if they know that there is a high probability that incorrect payments will be detected, that they will be required to repay any debt and that they may be prosecuted if they attempt, by fraud or misrepresentation, to obtain payments to which they are not entitled. While this program will increase levels of compliance,

2 In calculating total savings for a particular year overpayments are claimed at 70 per cent of the total detected overpayments to account for the difference between identified overpayments and overpayments raised by DSS: in many cases small overpayments are waived or written off as it is not economical to pursue recovery.

3 For 1991-92, Actual savings = reduction in fortnightly outlays *times* 7.6 (where 7.6 represents the number of fortnights occurring in the first year of the program's operation) *plus* 70 per cent of overpayments identified. For 1992-93 and 1993-94, Actual savings = reduction in fortnightly outlays *times* 13 (where 13 represents the number of fortnights occurring in the year) *plus* 70 per cent of overpayments identified *plus* the previous year's carryover.

4 For 1991-92, savings carryover represents 18.4 fortnights savings accruing in the first year of the project. For 1992-93 and 1993-94, it represents 13 fortnights' savings.

5 For 1993-94, Actual savings = reduction in fortnightly outlays *times* 9.75 (where 9.75 represents the number of fortnights occurring in the first year) *plus* 70 per cent of overpayments identified.

6 For 1993-94, savings carryover represents 16.25 (= 26 - 9.75) fortnights' savings accruing in the first year of the project.

quantifying this effect is difficult and no estimation of the financial benefit to the Commonwealth has been attempted.

Appendix C: Technical standards report

The purpose of the Technical Standards Report is:

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

As to the items listed in paragraph 58, the following guidance is provided.

Data Quality

As part of the Technical Standards Report the matching agency should, where practicable in consultation with the source agencies or organisations, compile a data dictionary for all data which is supplied as part of the program which includes the following:

- a file description of each file used by the matching agency which 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, eg, yy or yy mm dd, annual income, amount in thousands.

Matching techniques

The following matters should be clearly documented:

- the matching algorithm used. For example, first 6 characters of surname and value of forename, together with date of birth; phonex equivalent of family name and equality of birth year;
- the risk to be addressed;
- rules for recognising matches;
- the destination of the results of the match;
- 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 (e.g. annual income versus financial year income).

Data Quality Controls and Audit

The following should be documented:

- any relevant control measures used to ensure data quality, such as the timing of any extract files that may be taken for the matching program; and
- any audit programs to which the data has been, or is regularly, subjected;

Security and Confidentiality

Precautions proposed be taken at all stages to ensure that personal data used in and arising from a data-matching program:

- are not subject to accidental or intentional modification;
- are 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
- are not disclosed other than as intended by the program protocol.

Specific reference to access controls including password security, audit trail including logging of access and similar controls.

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 D: Statement of costs and benefits for data-matching programs

Introduction

The data-matching guidelines call for agencies to prepare statements of costs and benefits:

- when starting a new program - this statement would form part of the program protocol;
- when preparing a program protocol for a program which was already running when the matching agency adopted the guidelines; and
- when evaluating programs.

While it is clearly 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 which a formal cost-benefit analysis aims to produce will not be the most meaningful way of presenting the impact of the program. The degree of detail appropriate will vary depending on the nature of the program - for example, a

less detailed statement might be appropriate for a program whose function was to carry out a task which 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 could be used as a check-list by agencies, though noting that not all categories of costs and benefits will apply to any one program.

Consideration of these guidelines by the Audit Office

The Australian National Audit Office seeks to promote best practice in public administration. It has indicated that these guidelines are consistent with what it considers to be good practice in such cost benefit assessments. However, these comments from the ANAO do not prejudice the Auditor General's right to audit data-matching activities and to comment unfettered on any cost benefit assessments.

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 reasons for believing 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 program's performance.

A statement which 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.

It is most important that statements of estimated costs and benefits indicate clearly the method by which the estimates were reached, and any assumptions on which they were based.

Methods of presenting cost/benefit information

Perspective

It is suggested that statements of costs and benefits be presented from the perspective of the Commonwealth, rather than an 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 program would 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 programs which 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 to led to further investigation; and
  - the outcomes from investigation of cases.
- for new programs, any pilot project or other preliminary assessment of the program. If estimates of costs and benefits are based on results from pilot projects, 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 ways - cultural, economic, legal and so forth - that make parallels difficult to sustain. Such comparisons should be used, if at all, with caution 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 program and the information available. In all cases the statement should compare the outcomes from the data-matching program and the outcomes which would arise in the absence of the program.

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

- Compare the costs and benefits of the program with the most likely use of the resources involved, that is, what the 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 instance, 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 program against the costs and benefits of achieving the same result by some alternative method (for example, a manual process). This would be most appropriate where:
  - one of the main benefits of the program was administrative savings from efficient performance of a task which otherwise would have to be carried out by other means; or
  - the benefits of the program are quantifiable, but not financial (for example, detection and prosecution of people breaking a law). By comparing the cost of the data-matching, and the costs of achieving

similar results by other methods, a good measure of the effectiveness of data-matching can be obtained.

- If neither alternative uses of the resources required for the program nor the costs of achieving the same outcomes by alternative means can 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 patently impossible to apply either of the other approaches: it gives much less useful information about the program and makes it much more difficult to judge whether it constitutes an efficient application of resources. The reasons why neither of the other approaches could be used should be spelt out.

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 program.

This type of calculation is only regarded as necessary for data-matching programs if they 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 and the discount rate explicitly stated.

Information on calculating net present values can be found on page 9 of the Department of Finance 'Introduction to Cost-Benefit Analysis for Program Managers'.

Further information on cost/benefit analysis

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

- 'Guidelines for Costing Government Activities';
- 'Introduction to Cost-Benefit Analysis for Program Managers'; and
- 'Handbook of Cost-Benefit Analysis'.

These publications are available from the Australian Government Publishing Service.

Estimating Costs

The costs of programs can be broadly divided into:

- establishment costs, comprising
  - staff costs involved in setting up the program, for example, staff time to develop appropriate systems to process and handle the results;
  - capital costs; and

  - other costs, such as publicity costs and computer time; and

- running costs, including
  - costs associated with conducting runs of the program, for example the cost of maintaining the system; and
  - costs associated with following up matches, such as legal costs, the costs of appeals processes

For many programs, some of these cost categories may be negligible (for example, many programs do not involve capital expenditure) and can be ignored.

Establishment costs

Staff costs. A suggested approach to estimating the cost of staff time involved in developing a new data-matching project is:

- Estimate the amount of time (in person weeks, months or years) that will be spent by all staff on initial development of the project.
- Estimate the average salary of project staff (per week, month or year), and multiply it by the estimated staff time needed for project development.
- Multiply the result of step 2 by a factor to allow for labour on-costs and overheads. The Department of Finance recommends multiplying the basic salary cost by 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 "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 rule of thumb, it will often 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.

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 they are negligible and have been left out should be included in the protocol.

Costs associated with following up matches. The costs associated with taking action on the strength of the output from the program will also tend, predominantly, to be staff costs. The method used to estimate these costs will depend on the way the program results are used.

- If follow-up of data-matching results is carried out by dedicated staff, the cost of this activity can be readily calculated.
- If staff who carry out reviews based on data-matching results also have other functions, the time required for those reviews could be calculated either by estimating the proportion of time which 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 following up data-matching results is decentralised, and data on how review staff apportion their time is not available.

Estimating Benefits

The benefits of programs can be broadly divided into:

- direct financial benefits, including
  - recovery of incorrect payments;
  - prevention of incorrect payments; and
  - increased revenue collection;
- indirect financial benefits, such as
  - administrative savings;
  - benefits of voluntary compliance (deterrence effects);
- 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, either because some amounts are too small to warrant recovery action, or because recovery action is unsuccessful; and
- the cost of recovery action (unless this is explicitly included under the costs of the program).

As an example, one agency has estimated that 70 per cent of overpayments identified represent actual savings, ie, 30 per cent of overpayments cannot be recovered or will be absorbed in recovery costs.

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 amounts actually recovered and the cost 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 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 be sure how much would have been incorrectly paid had the payment not been terminated or reduced.

If a rule of thumb is used (for example that the incorrect amount would have continued to be paid for a standard period) the rule should be specified, and the reasons for adopting it made explicit (for example, an agency might adopt a rule that an incorrect payment would have continued to be paid for half the average period for which payments of that type are made).

In adopting this sort of rule, agencies should take account of the possibility that other review methods 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 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, which would in any case have to be performed, to be carried out 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 headed "Methods of presenting information").

Voluntary compliance. Agencies may think that public knowledge that a program is operating leads to increased compliance with the law. Benefits of this sort are 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 which 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 programs anticipated impact on them.

Non-financial benefits

Many programs have benefits which 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 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 program is considered worthwhile, and provides a basis for comparing actual performance against initial estimates.

If a 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 programs which have quantifiable non-financial benefits are presented above under the heading Methods of presenting information.