Data Matching in the Public Interest

A guide for the Victorian public sector

Edition 1 – August 2009

privacy victoria

Office of the
Victorian Privacy
Commissioner

COVER PHOTO: www.istockphoto.com

Table of Contents

1. Introduction

Technology has made data matching on a large scale between organisations quicker and easier. Parliament recognised the significance of this when it gave the Privacy Commissioner a function 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 on personal privacy are minimised, and to report to the Minister the results of the research and monitoring*[1]

The key role of state and local government in providing services across a range of areas necessitates the collection, storage, use and disclosure of vast amounts of personal information. The desire to deliver better and more efficient services has seen a move to 'seamless' or 'joined up' government models, and thus data matching plays an increasingly important role.

Data matching has significant potential benefits in many varied contexts, in particular as technology improves. It may assist law enforcement, revenue protection, research, more efficient use of resources and better delivery of services. The focus in this guide's title on data matching "in the public interest" is intended to acknowledge that privacy is but one of many public interests that are relevant to data matching activities.

Depending on how it is conducted, data matching also poses privacy risks, including concerns about legitimacy and community expectations, secondary use, function creep, data quality, automated decision-making, constructive identification and profiling.

Even where the proposed data matching would clearly be lawful under the *Information Privacy Act 2000 (Vic)*, it is still essential for any organisation proposing to match personal information to consider whether the information should be matched and the reasons for it. Any data matching will involve inherent risks to privacy and these need to be balanced against the potential positive outcomes to be achieved. Data matching needs to be authorised, transparent and conducted according to appropriate standards in an accountable manner. When this happens, the legitimacy of specific data matching exercises is enhanced and their benefits increase.

This guide aims to help Victorian public sector organisations to assess existing or contemplated data matching activities for compliance with the *Information Privacy Act*, as well as consistency with the *Charter of Human Rights and Responsibilities* (the Charter) and good privacy practice.

This guide was prepared with the assistance of Anna Johnston of Salinger Privacy and Anthony Bendall, Deputy Privacy Commissioner. My thanks also to Privacy Victoria staff, past and present, who assisted with consultation workshops.

I hope that all organisations will find this guide useful. Our staff are always available to advise those undertaking or considering data matching activities.

Helen Versey
Privacy Commissioner

[1] *Information Privacy Act 2000 (Vic)* Section 58(m).

2. Scope of this guide

Aim of this guide

This guide has been developed to meet a need of the Victorian public sector, including Local Councils.

Data matching is an activity which creates special risks and considerations for privacy. Privacy Victoria's 2005 audit of data matching activities found that most of the organisations surveyed have difficulty in translating a high level of awareness of their privacy responsibilities into an understanding about how the Information Privacy Principles (the IPPs) apply to their data matching practices. Most organisations favoured the development of data matching guidelines from the office to assist in this regard.[2]

Privacy Victoria has therefore developed this guide to assist organisations to assess existing or contemplated data matching activities for compliance with the Information Privacy Principles in the *Information Privacy Act*, as well as consistency with the Charter of Human Rights and Responsibilities and good privacy practice.

Data matching is an activity which creates special risks and considerations for privacy.

The guide is not intended to cover the *Health Records Act 2001 (Vic)*. If your organisation handles health information, you will need to comply with the Health Privacy Principles (HPPs) in that Act as well. Advice on compliance with the HPPs should be sought, in the first instance, from the Office of the Health Services Commissioner, which regulates the handling of health information in Victoria.

This guide is not intended to deal with issues concerning ownership of information or copyright. Responsibilities of participant organisations are discussed in light of privacy and data protection requirements, which apply regardless of who 'owns' the personal information involved.

This guide is not legally binding and does not constitute legal advice about how organisations are to comply with the IPPs in specific circumstances. It is intended to indicate how the Privacy Commissioner interprets and applies the IPPs.[3] This guide indicates the matters the Privacy Commissioner may consider if organisations seek advice about data matching, when dealing with complaints about data matching, or when conducting an investigation into an apparent breach of the IPPs involving data matching.

The details of compliance with the IPPs are ultimately up to each organisation. Organisations should consult their privacy officer or unit as required, and may wish to seek independent legal advice where appropriate. Officers from Privacy Victoria are available to take enquiries and provide guidance by telephone on 1300 666 444 or by email at enquiries@privacy.vic.gov.au.

---

[2] Privacy Victoria, *Victorian Public Sector Data Matching Audit*, February 2005.

[3] Section 58 of the *Information Privacy Act* sets out the functions of the Privacy Commissioner, which include "(a) to promote an understanding and acceptance of the Information Privacy Principles and of the objects of those Principles"; and "(s) to provide advice (with or without a request) to any individual or organisation on any matter relevant to the operation of this Act".

In the public interest

The focus in this guide's title on data matching "in the public interest" is intended to acknowledge that privacy is one of many public interests that are relevant to data matching activities. There are clearly many legitimate reasons for engaging in data matching, and these are generally acknowledged in the permissible categories for use and disclosure set out in IPP 2. The public interest in law enforcement, protecting the public revenue, carrying out research, and protecting the public's health, safety and welfare will often justify data matching activities.

The public interest in law enforcement, protecting the public revenue, carrying out research, and protecting the public's health, safety and welfare will often justify data matching activities.

However ensuring safeguards are in place to maintain public confidence in the responsible and transparent handling of personal information is also in the public interest.

This guide aims to help Victorian public sector organisations find the appropriate balance between privacy and other public interests when engaged in data matching.

However ensuring safeguards are in place to maintain public confidence in the responsible and transparent handling of personal information is also in the public interest.

3. What is data matching?

A definition

Although in common use, the term 'data matching' means different things to different people. No standard definition exists. Various activities involving the comparison of separate data sets may be referred to as 'data matching'.

The terms 'data matching', 'data linking' and 'data cleansing' are sometimes used interchangeably with 'data matching', although data linking and data cleansing are actually specific types of data matching activities. Data cleansing refers to intra- or inter-organisation comparisons of personal information designed to improve data quality, while data linkage refers to linking data sets to produce new kinds of information about individuals in either an identified, de-identified or 'anonymised' form.

For the purposes of this guide we have continued with the definition developed for this Office's 2005 audit of data matching activities:

> *Data matching includes any/all of the following:*
>
> - *comparing personal information from two or more records to determine whether personal information from different records matches to the same individual; and/or*
> - *comparing personal information about an individual obtained from two or more records to gauge the accuracy of the personal information about that individual in each of the records, and to improve the accuracy of the personal information in all of the records; and/or*
> - *connecting two or more records of personal information to aggregate personal information about an individual.*

Purposes of data matching

In government contexts, data matching is often used to compare separate sets of personal information in order to assess or confirm specific persons' rights, benefits or privileges. Data matching for determining eligibility for benefits may include transport authorities seeking to validate a claim to a concessional fare by collecting data from Centrelink; or a benefits organisation checking with the Registry of Births Deaths and Marriages that an application has not been lodged in the name of a recently deceased person.

However data matching may also be intended for research purposes, such as studies of public health trends. Data matching for research purposes might, for example, provide bulk data for epidemiological research so that public health researchers can build up a picture of what is normal and abnormal across a population.

Data matching may also be used for the purposes of law enforcement, including the protection of the public revenue, such as collecting ratepayer data from Local Councils to enable fines enforcement. Preparing for or responding to an emergency (e.g. Local Councils and the Country Fire Authority matching data to combat fire hazards) is another possible purpose of data matching.

Data matching can also facilitate consumer profiling and other commercial uses of personal information. These activities may not always affect individuals' rights, benefits or obligations in their relationship with Victorian government organisations, but they may be using information obtained by those government organisations.

Privacy Victoria's 'snapshot' audit of data matching activities in the Victorian public sector found that there was a fairly even distribution of the purposes for which data matching is used.[4] Responses listed law enforcement (47%), research (40%), and protection of the public revenue (47%) as the main purposes for data matching. Most surveyed (73%) also selected 'other' as the main purpose for data matching. Examples of other purposes included data cleansing, verifying a birth (e.g. matching with the Registry of Births, Deaths and Marriages) or updating a record (e.g. matching property details with a ratepayer).

The most likely 'public interest' purposes for which data matching will be sought are:

- improved efficiency or service-delivery
- research
- law enforcement
- protection of the public revenue, or
- preparing for or responding to an emergency.

Each of these five purposes creates specific considerations because each relies on a different permitted disclosure under IPP 2; later in this guide we have elaborated on these purpose-based considerations.

Who is involved?

Slightly different considerations will arise, depending on with whom your organisation intends to conduct data matching.

DATA MATCHING WITHIN A VICTORIAN PUBLIC SECTOR ORGANISATION

Internal data matching would normally be considered 'use', rather than 'disclosure', although this is not an absolute rule, particularly in the case of large public sector organisations, with numerous, discrete business units.[5] In any event, IPP 2 needs to be complied with for either use or disclosure. Slightly different considerations might apply, though: for example, individuals may have clearer expectations about matching within an organisation. In addition, data security might present less of a problem because the information is staying within the custody and control of one organisation.

---

4 Privacy Victoria, *Victorian Public Sector Data Matching Audit*, February 2005.

5 *KJ v Wentworth Area Health Service* [2004] NSWADT 84.

DATA MATCHING BETWEEN TWO OR MORE VICTORIAN PUBLIC SECTOR ORGANISATIONS

Data matching between two or more organisations will amount to both 'disclosure' and 'collection'. This may mean that individuals are less likely to expect data matching to occur, making it more difficult to rely on IPP 2.1(a)(ii), which allows disclosures related to[6] the primary purpose of collection only if the individual would reasonably expect it. Thus in most cases, data matching should be conducted either for the primary purpose of collection, with consent, under legislated authority, or for one of the 'public interest' purposes allowed under IPP 2.1(c) to (h). Each of the public interest purposes are discussed later in this guide.

VICTORIAN PUBLIC SECTOR ORGANISATION MATCHING WITH PRIVATE SECTOR

Data matching with the private sector will also clearly constitute 'disclosure' for the purposes of IPP 2. It is also far less likely that data matching for a related secondary purpose would be expected by the individual. Moreover, as most private sector organisations are not subject to the *Information Privacy Act*, protections relating to the information once disclosed may vary enormously.

While some private sector companies and organisations are subject to the National Privacy Principles (NPPs) in the *Privacy Act 1988 (Cth)*, many are not.[7] Even where the private sector body is subject to the NPPs, the protection offered by the NPPs is, in some instances, less stringent than the IPPs. This is one of the factors to be considered by a public sector organisation before deciding to engage in data matching with the private sector. Moreover, as many private sector companies operate across state and territory borders, issues under IPP 9 (Transborder Data Flows) may arise.[8]

VICTORIAN PUBLIC SECTOR ORGANISATION MATCHING WITH COMMONWEALTH OR INTERSTATE PUBLIC SECTOR

Matching data with government agencies outside Victoria raises issues that are similar to when matching information with the private sector. Privacy protections vary between jurisdictions: while NSW, Queensland, Tasmania, the NT and ACT have privacy legislation, the principles that apply under those statutes differ from the IPPs. Likewise, while Commonwealth agencies are bound by the Information Privacy Principles in the *Privacy Act*, these also differ significantly from the Victorian IPPs. At the time of writing South Australia and Western Australia have no legislative privacy protections, although South Australia operates under an administrative regime.[9]

There are also clear implications under IPP 9 (Transborder Data Flows), especially where no privacy law will apply to the receiving organisation. In that situation, your organisation needs to take steps to ensure that privacy protections are built into the contractual agreement under which information is to be matched.[10]

---

6 If the personal information is "sensitive information", the secondary purpose must be *directly* related to the primary purpose of collection.

7 Particularly if they are "small" businesses, with an annual turnover of less than $3 million per annum; see *Privacy Act,* s.6D.

8 See discussion above and below.

9 See Privacy Victoria, *Privacy Regulation across Australia, as at 1 October 2008*.

10 See Privacy Victoria, *Model Terms for Transborder Data Flows of Personal Information*. June 2006.

Data linkage

Data linkage is a particular type of data matching, involving the linking or aggregation of different data sets. Data linkage often involves the use of a Statistical Linkage Key (SLK). An SLK was defined in Privacy Victoria's audit of data matching activities in the Victorian public sector as:

> *"a derived item based on an assortment of information from several other data items. An example of a constructed statistical information key is the use of the 2nd 3rd and 5th letters of a person's first name, the 2nd, 3rd and 5th letters of a person's surname and the person's date of birth".*[11]

SLKs are often argued to be privacy enhancing, although in some instances, SLKs can be linked back to a particular individual, meaning that the information generated remains "personal information" and raises privacy issues.

In some cases, using a code or stripping identifiers from the data may not be sufficient to de-identify a person to third parties, or to render the information anonymous or unidentifiable (and thereby removing it from the reach of the *Information Privacy Act*). For example, an SLK comprised of an individual's initials and date of birth may not adequately anonymise a person, especially if the key can be 'unlocked' by anyone holding these details–which are routinely provided by individuals, for example, to access government services or benefits.

A scoping audit conducted by this Office in February 2005 found that SLKs were used by one out of every four government agencies surveyed.[12] Organisations should ensure that linkage keys are not purported to be used to anonymise data where re-identification is reasonably possible. A similar caution has been asserted by an Australian Ministerial Council working group in its work on statistical data linkage:

> *It is a very common misconception that an SLK [statistical linkage key] by itself does not allow an individual to be identified when attached to non-identifiable data….*
>
> *For example, a common SLK (for example, the [joint Commonwealth/State/Territory Home and Community Care program (HACC)] linkage key) includes the gender and date of birth plus three characters from known positions within the surname and a further two from known positions within the first given name. Information at this level contains much that could be used to identify individuals.*
>
> *The HACC linkage key is primarily a tool used to uniquely identify an individual with a high degree of reliability, without regard for that individual's identity. It is not a tool primarily designed to protect or ensure the anonymity of the individual. While the HACC linkage key does provide some protection to clients to ensure that clients are not unintentionally identified, it is not (on its own) sufficient to provide complete privacy protection. Thus, the key issue for the HACC and other existing linkage keys has been to identify and develop other appropriate safeguards and measures (for example, protocols, encryption processes) to ensure that these requirements are met in conjunction with the use of the SLK….*[13]

---

[11] Privacy Victoria, *Victorian Public Sector Data Matching Audit*, February 2005.

[12] Office of the Victorian Privacy Commissioner, *Victorian Public Sector Data Matching Audit*, Audit 01.05, February 2005, page 3,.

[13] Australia, Community Services Ministers' Advisory Council, National Community Services Information Management Group, Statistical Linkage Key Working Group, *Statistical Data Linkage in Community Services Data Collections*, May 2004, see especially pages 12-13, http://www.aihw.gov.au/publications/hwi/sdlcsdc/sdlcsdc.pdf.

Risks posed by data matching

INHERENT RISKS

Data matching between different organisations will almost always be an example of function creep, because it features secondary use and indirect collection.

Data will typically be disclosed by the source organisation for reasons not originally intended or explained to the subject at the time of original collection, as well as collected by the recipient organisation other than directly from the subject individual.

Data matching therefore poses an inherent interference with privacy, as it is contrary to the basic 'use limitation' privacy principle that information should only be used for the original purpose of collection. As noted elsewhere in this guide, this interference with privacy must therefore be justified by a countervailing 'public interest'.

The very nature of matching data also raises data quality and transparency risks. When the data is being transmitted across two or more organisations, data security risks are also heightened. Data matching for law enforcement purposes may also subvert our criminal justice principle of 'innocent until proven guilty'.

The creation of a new dataset from the results of a data matching exercise also generates its own risks of excessive or intrusive collection, excessive data retention, data security (including unauthorised access or disclosure), and further function creep if the new dataset is later used or disclosed for purposes beyond that approved for the data matching program.

Data matching therefore poses an inherent interference with privacy, as it is contrary to the basic 'use limitation' privacy principle that information should only be used for the original purpose of collection.

HARM TO THE SUBJECT INDIVIDUAL

The spread of data through inappropriate data matching can lead to an increased risk of identity fraud or identity theft, which directly harms individuals.

Data matching may also lead to unjust decisions or outcomes for individuals if they are denied a benefit or targeted for investigation based on incorrect data about a person's identity or circumstances, or if information is taken out of context.

Automated decision-making, in particular, heightens these risks as "unrelated, unstable, outdated or unreliable data … (enters) the decision-making pathway without the data flaw being identified or critically examined by an officer".[14] Decision makers at a recipient organisation may not be apprised of the level of reliability of the data that has come from data matching rather than from their own records.

[14] Australian Government Information Management Office, *Automated Assistance in Administrative Decision-Making: Better Practice Guide*, February 2007, p.31.

Individuals may also be harmed or embarrassed if their personal information is constructively identified through a data matching exercise. Constructive identification is more likely when datasets are small, or when only a small number of individuals within a dataset have particular characteristics.

For example, indigenous status and educational outcomes, if reported at a school level, would enable constructive identification if there were only one or two indigenous students in a class year. Data about deaths and causes of death can also easily lead to constructive identification because of the publicity surrounding certain types of death, such as homicide, road accidents and child deaths.

DATA QUALITY CONCERNS

Data quality problems can arise when systems design has been too inflexible to accept blank entries. For example, if date of birth is set as a mandatory field for one dataset, data entry operators will insert dummy data if they do not know the correct date of birth. This will cause problems for data matching with other datasets which either rely on, or seek to validate, an individual's date of birth.

Data about ethnicity and indigenous status may also cause difficulties because of the voluntary and self-identifying nature of most reporting.

Inaccuracies can also arise because different organisations have different classifications for the same data, or because individuals have (legitimately) provided different data to different organisations. The source datasets are likely to have been collected at different times and for different purposes, leading an individual to provide different, but equally truthful, answers to requests for data such as their surname (maiden name, married name or other name changes), first name (Alex, Alexander or Sandy), address (postal, street, home, business, holiday house, as well as address changes), and so on[15]. Different organisational rules may also cause people to describe their occupation, income or assets differently over time.

AGGREGATION AND PROFILING

Unique identifiers can be critical to data matching. If you know that a person has the same unique identifier in the database of two or more organisations, it becomes much easier to match the different data held about the person in each of those organisations.

One of the reasons any proposal for a universal ID card is sensitive, from a privacy point of view, is that such a system would ordinarily assign a unique identifier to every person, and with this identifier it would be possible to bring together many different types of information about the person. Depending on the extent and sensitivity of the connected information, a profile of a person could emerge that is more detailed than would be legitimate for any government to compile, especially if various data sets from the public and private sectors were connected using the unique identifier.

This is why privacy and data protection laws pay strict attention to unique identifiers. It does not mean that unique identifiers cannot be created and used for proper purposes, including properly authorised and transparent data matches.

[15] Particular difficulties can arise with structure of ethnic names and persons using a "preferred name".

4. The legal environment affecting data matching

Compliance with the IPPs

Unlike some other jurisdictions such as New Zealand,[16] there are no provisions in the *Information Privacy Act* that specifically deal with data matching activities. Rather, when public sector organisations wish to match data, they need to do so in such a way as to comply with the Information Privacy Act and the IPPs.

The *Information Privacy Act* only applies to "personal information", which is recorded information or opinion, whether true or not, about an identifiable living individual.[17] Personal information can be almost any information linked to an individual, including name, address, sex, age, financial details, marital status, education, criminal record or employment history. However, information about an individual's health, and information collected, used or disclosed in the context of providing a health service, is specifically excluded from the coverage of the *Information Privacy Act* and is instead regulated by the *Health Records Act*.

The IPPs may affect an organisation's data matching activities in a variety of ways. For instance, where an organisation wants to match data that it holds with data held by another organisation, it will be planning to collect personal information from that other organisation, so IPP 1 (Collection) must be considered.

To match data is to use it; to provide data for matching by another organisation is to disclose it; and so in both instances IPP 2 (Use and Disclosure) is relevant.

To match data is to use it; to provide data for matching by another organisation is to disclose it; and so in both instances IPP 2 (Use and Disclosure) is relevant.

Where a match combines data sets to produce new information about a person, he or she has a right of access to it and a right to seek its correction (*Freedom of Information Act 1982 (Vic)* and IPP 6). And data matching is quintessentially an area in which the issue of unique identifiers arise (IPP 7). Potentially IPP 9 (Transborder Data Flows) and IPP 10 (Sensitive Information) will also apply.

Consistency with the Charter

In addition to the IPPs, the privacy rights of Victorians are protected under the *Charter of Human Rights and Responsibilities*.[18] The Charter itself expressly provides that the human rights it protects may be subject, under law, to some reasonable limits. But those limits must be "justified in a free and democratic society based on human dignity, equality and freedom".[19]

---

16 *Privacy Act 1993 (NZ)*, Part 10 (sections 97 to 109).

17 Section 3 of the *Information Privacy Act 2000 (Vic)*; see Privacy Victoria, *Guidelines to the Information Privacy Principles*, (Guidelines to IPPs), September 2006, pp. 7 to 14.

18 *Charter of Human Rights and Responsibilities Act 2006 (Vic)*

19 Section 7(2), *Charter of Human Rights and Responsibilities Act*

The Charter does not provide any new avenue of redress for individuals who believe their privacy has been breached. But it does impose an obligation on all public authorities to act in a way that is compatible with the human rights protected by the Charter.[20]

The 'privacy' protected by the Charter is a broader concept than just the privacy of personal information, regulated by the IPPs. The dimensions of privacy that need to be considered include:

- bodily privacy (to protect the integrity of the physical person);
- territorial privacy (to protect personal space, objects and behaviour);
- communications privacy (to protect against eavesdropping);
- locational privacy (to protect against surveillance); and
- information privacy (to protect personal information).

Data matching activities for the most part will only involve information privacy. However if a data matching activity draws upon records created from surveillance, DNA samples or monitoring of location or communications , these other elements of privacy will also need to be considered with respect to the Charter right to privacy.

The right to privacy protected by the Charter is expressed as the right of a person "not to have his or her privacy, family home or correspondence unlawfully or arbitrarily interfered with".[21] The right to privacy under the Charter is thus qualified by an interference having to be "unlawful" or "arbitrary".

Even if data matching is authorised under statute or is otherwise lawful, it may amount to a breach of the Charter if any resultant interference with privacy could be considered arbitrary. The concept of arbitrariness means that even interference provided for by law should be in accordance with the aims and objectives of the Charter and be reasonable in the circumstances.[22]

In the context of data matching, this suggests that any impact on individuals must be fair. Protecting and enhancing data quality and integrity will play a key role in ensuring that the impact on individuals affected by data matching is not unfair.

Other legislation

Some data matching is authorised by statutory provisions apart from the *Information Privacy Act*. Equally, many public sector organisations are regulated by organisation-specific pieces of legislation that impose strict confidentiality requirements. Those obligations override the provisions of the *Information Privacy Act* where there is inconsistency. So although the *Information Privacy Act* may permit the disclosure inherent in a particular data matching exercise, the organisation-specific legislation may not.

---

[20] Section 38, *Charter of Human Rights and Responsibilities Act*

[21] Section 13, *Charter of Human Rights and Responsibilities Act*

[22] The concept of arbitrariness versus lawfulness is illustrated in a decision of the United Nations Human Rights Committee, *Toonen v. Australia*, Communication No. 488/1992, U.N. Doc CCPR/C/50/D/488/1992 (1994); available at www1.umn.edu/humanrts/undocs/html/vws488.htm, last accessed June 2009.

5. Considerations when contemplating data matching

Are you disclosing or collecting?

When first contemplating a data matching program, it is important to be clear about what role each party will be playing. Will your organisation be providing source data for the program? Will your organisation be receiving some kind of 'output' or results from the program?

In short, will your organisation be disclosing personal information, or collecting it–or both?

Different obligations will apply, depending on which role/s your organisation will play in the data matching program.

When first contemplating a data matching program, it is important to be clear about what role each party will be playing.

A disclosing organisation is effectively the source of data to be used in the data matching program. The disclosing party must be particularly wary of compliance with its disclosure obligations under IPPs 2, 7 and 9 (see more about these below), as well as other legislative or contractual limitations on disclosure, such as secrecy provisions or confidentiality agreements.

The disclosing or 'source' organisation best understands the data's strengths and weaknesses. It is in the best position to determine whether the data will even assist a recipient before it is provided. The disclosing organisation can also advise on data quality risks specific to that dataset.

The disclosing organisation is also best placed to ensure transparency of the data matching activity, by providing notice to the affected subjects of the data. For this reason compliance with IPP 1.3(d)–notice as to the types of organisations to which the organisation usually discloses information of that kind–is the responsibility of the disclosing organisation.

A collecting organisation will be the recipient of results of the data matching program. As such, they must be particularly concerned with compliance with collection-related obligations under IPPs 1, 7, 8 and 10 (see more about these below). Subsequent use of the results must also comply with IPP 2, and the data must be appropriately secured in accordance with IPP 4.

The flowchart at Figure 1 (page 22) outlines the steps to take when contemplating a data matching program, and illustrates whether each step is common to both disclosing and collecting organisations, or poses different considerations for each distinct role. Later flowcharts, outlining the steps to take during the Design and Operational phases, also illustrate the different considerations for each party.

Of course many data matching programs will feature two or more organisations effectively 'pooling' their data, in order to create a new, value-added data set for use by each of the participating organisations. In such cases, an organisation may be in the role of both disclosing their source data, **and** collecting the results of the data matching. Organisations in this position must therefore comply with **both** sets of obligations.

The benefits of a Privacy Impact Assessment

Privacy Impact Assessment (PIA) has been defined as "an assessment of any actual or potential effects that the activity or proposal may have on individual privacy and the ways in which any adverse effects may be mitigated".[23]

A PIA considers the future consequences of a current or proposed action, and looks to prevent or minimise any negative impacts on privacy. This involves assessing any benefits that the data matching could bring to society or individuals and balancing them against any negative effects, such as erosion of privacy rights or the likelihood of damage, distress or embarrassment being caused to individuals. A PIA is an important tool to avoid or minimise the risk of harm. When conducted in consultation with stakeholders, a PIA can also test community expectations and engender public trust in a proposal.

While PIAs are not formally required under the *Information Privacy Act*, most Victorian public sector organisations are now required to comply with the Victorian Government Risk Management Framework.[24] In order to do so, all risks arising from a project, including risks to privacy, need to be identified and addressed. A PIA should review the matters identified below, amongst others, as part of a sensible risk management approach.

Further guidance on how to conduct a PIA is available in Privacy Victoria, *Privacy Impact Assessments: A guide for the Victorian Public Sector*, Edition 2, April 2009.

Legitimacy of project objectives

The underlying rule of IPP 2 is that information supplied for one purpose should not be used for another unrelated purpose, unless specific exceptions apply. Data matching has the potential to undermine that principle. Organisations may find it very useful to match, say, property ownership data with drivers' licence data or information from the electoral roll. But from the point of view of the public, these are three very different data sets in which their personal information is held (often compulsorily) to serve quite separate public purposes.

At a minimum, when unrelated data sets are connected and compared–especially for reasons that may affect persons' legal rights and obligations–that activity needs to be:

- clearly justified in the public interest;
- authorised under law;
- carried out transparently; and
- supervised independently.

---

[23] Blair Stewart, "Privacy Impact Assessments", (1996) 3 *Privacy Law + Policy Reporter* 61, 62.

[24] Local Councils are not covered by the Framework as at the date of writing. The Framework is available at the Victorian Managed Insurance Authority (VMIA) website, www.vmia.vic.gov.au which also contains other risk management tools and advice.

Before any data matching occurs, the participating organisations should agree and document the objective that the proposal is designed to achieve. The objective of the data matching proposal must be a legitimate objective having regard to the functions of the organisation, including any functions conferred by legislation.

The objective of the proposal must also accord with one or more of the 'public interest' activities envisaged by the IPPs, such as law enforcement, protecting the public revenue, carrying out research, or protecting the public's health, safety and welfare, unless expressly authorised or required by another law.

Only once this is done can further privacy issues be addressed, such as deciding whether identifiable personal information needs to be matched, or whether anonymised or purely statistical information would be sufficient.

Project management considerations

For any project, an organisation must ensure its proposal will be lawful, a proportionate response to a public interest problem, and an effective response to that problem.

LAWFULNESS

- Is there power to carry out the function to which the data matching relates? Is it express or implied?
- Organisations need to be clear about whether legislation that regulates them permits the data matching they engage in. Previous consultation work of Privacy Victoria has revealed that this is not always the case. Where there is general legislative authority, the organisation may still need to consider its parameters. For example, if an Act allows disclosure to occur where it is "reasonably necessary", the organisation needs to consider what is meant by that term and whether it only authorises disclosure of individual records on a case by case basis, or in fact sanctions an entire data matching regime.
- Should specific legislative authority be created to permit the data matching?
- Are there express statutory restrictions on the data matching activity proposed, or any restrictions which may be implied by other statutory provisions or common (judge made) law?

PROPORTIONALITY

- Is the public interest in allowing the data matching sufficient and compelling?
- Does the program involve data matching on a scale that is excessive, having regard to:
  - the likely benefit to be derived;
  - the number of organisations that may be involved; and
  - the amount of personal information to be matched?

- Is it more appropriate for the source agency to retain its personal information, and to perform matching activities on behalf of the organisation that seeks to use or benefit from it? This is related to the 'best custodian' principle,[25] which means that the best custodian of a large set of personal data is usually its original collector. The best custodian should retain control of the data, maintain its quality and manage any necessary and authorised access to it by others. However in some circumstances matching activities performed by the original custodian may not be appropriate, as it involves collection of unnecessary and potentially damaging information, such as when criminal activity is being investigated.

EFFECTIVENESS

- Conduct a cost/benefit analysis of the project. Given the costs of the proposed data matching, will the data matching program achieve monetary savings that are both significant and quantifiable, or other comparable benefits to the public?
- One way to measure this might be to consider conducting a smaller scale pilot project. However, even a pilot project should only be conducted after proper consideration of the factors set out above and below, and after conducting or commissioning a PIA.

General privacy considerations

Further considerations for any data matching proposal should include:

- Could the data matching program result in arbitrary interferences with individuals' privacy, such as unfairness or unforeseen harm?
- Could the subjects' consent be sought for the disclosures and collections to make up the data matching program?
- Could a data verification program (in which assertions are tested and provided with a yes/no response for each data field tested) be used instead of more detailed data matching to create new or enriched datasets?
- Are there less privacy-invasive methods of achieving the same public interest objective?

[25] See Privacy Victoria, *Large state datasets of personal information in the context of identity management and a federal government smartcard*, An address by Paul Chadwick, Victorian Privacy Commissioner, to the Public Record Office Victoria, 25 May 2006, pp.5-8.

Compliance with the IPPs

As noted above, different IPPs will apply, depending on whether your organisation will be disclosing personal information, or collecting it–or both–in the proposed data matching program. The discussion below clarifies which privacy principles are of most concern to disclosing (source) organisations, and which are particularly relevant to collecting (recipient) organisations.

Disclosing organisation

HOW SOURCE INFORMATION WAS COLLECTED

In the data matching context, the extent to which data matching can occur lawfully is largely determined by the way in which collection was undertaken by the source. Collection needs to have been: necessary,[26] lawful, fair, not unreasonably intrusive,[27] done with adequate and timely notice,[28] and where lawful and practicable, direct from the individual.[29]

Particularly important is the extent and type of notice given to the individual about the purposes for which their personal information was collected, and its anticipated use or disclosure. If data matching is or is planned to be a routine or regular occurrence, IPP 1.3(c) and (d) require the disclosing organisation to have given reasonable notice, at the time that it collected the information from the individual, that such data matching was one of the purposes for collection. The notice should include the names or types of organisations with which information is usually matched. Organisations should regularly check their privacy notice to ensure that it describes any organisations with which information is being matched.

The purpose for which information is compulsorily obtained will necessarily limit the extent and purpose for which that information can lawfully be used and disclosed.[30] In general, where an organisation has statutory powers to compel the provision of information to it, it may be considered an 'unfair' collection if information compulsorily obtained is subsequently disclosed for an unrelated purpose.

A collection of personal information may also be unfair if your privacy notice has misrepresented what will be done with the information once collected, such as claiming the information will be treated securely and confidentially when it is intended or proposed that the information be passed on to others for the purposes of data matching.

NECESSITY OF DISCLOSURE

IPP 2 sets out a general prohibition on the disclosure of personal information for unrelated secondary purposes, although there are various exceptions to that rule, when the public interest in the disclosure is seen to outweigh the public interest in the protection of privacy.

---

[26] IPP 1.1; see *Guidelines to IPPs*, pp. 21-22, 28-29.

[27] IPP 1.2; see *Guidelines to IPPs*, pp. 30-33.

[28] IPP 1.3, 1.5; see *Guidelines to IPPs*, pp. 33-39.

[29] IPP 1.4; see *Guidelines to IPPs*, p. 38.

[30] *Johns v Australian Securities Commission* [1993] HCA 56.

Depending on the public interest applicable to your data matching proposal, you may need to demonstrate that the disclosure of personal information from your organisation is "necessary". For example, IPP 2.1(c) allows disclosures if the disclosure "is necessary for research, or the compilation or analysis of statistics, in the public interest". Likewise IPP 2.1(d)(ii) allows disclosures if the organisation "reasonably believes that the … disclosure is necessary to lessen or prevent … a serious threat to public health, public safety, or public welfare".

Regardless of which public interest purpose applies, when disclosing personal information, the amount of information disclosed should not exceed what is sufficient to satisfy that purpose.

"Necessary" requires more than what may be administratively convenient or desired, but does not mean "essential". In *Ng v Department of Education*,[31] the Victorian Civil and Administrative Tribunal accepted that "necessary" means "subjected to the top scale of reasonableness". In 2004, the High Court of Australia reaffirmed that "necessary" does not mean unavoidable, essential or indispensable. The court suggested that an assessment of necessity must involve consideration of what is proportionate,[32] or "close scrutiny, congruent with a search for 'compelling justification'".[33]

Regardless of which public interest purpose applies, when disclosing personal information, the amount of information disclosed should not exceed what is sufficient to satisfy that purpose.

Further considerations, specific to the type of public interest purpose for which you propose to disclose personal information, are set out on pages 23 to 29 of this guide.

LAWFULNESS OF DISCLOSURE

Organisations need to be clear about whether legislation that regulates them permits the data matching they engage in. For example, although IPP 2 may permit disclosure of personal information to another organisation, confidentiality provisions in other legislation may prevent it.

IDENTIFIERS

IPP 7 deals with the assignment, adoption, use and matching of unique identifiers, such as driver's licence numbers, or unique patient identifiers. The Unique Identifiers principle is expressly intended to act as a safeguard against pervasive data matching across government.[34]

If your organisation proposes as part of your data matching program to disclose unique identifiers assigned to the subject individuals by another organisation, you must ensure you can comply with IPP 7.3. Such disclosures will only be allowed if:

a) the disclosure is necessary for your organisation to fulfil its obligations to the organisation which assigned the identifiers in the first place; or

---

31 *Ng v Department of Education* [2005] VCAT 1054 at para 77.

32 For further elaboration on the proportionality test, *Mulholland v Australian Electoral Commission* [2004] HCA 41 per see Gleeson CJ at paras 33-39, and Kirby J at paras 249-251.

33 *Mulholland v Australian Electoral Commission* [2004] HCA 41 per Gleeson CJ at paras 39-40.

34 *Guidelines to the IPPs*, pp. 135-144.

b) one or more of the permitted disclosures in IPP 2.1(d) to 2.1(g) applies (emergencies, investigations, authorised by another law, or law enforcement activities); or

c) you have obtained the consent of the individual to the disclosure.

TRANSBORDER DISCLOSURES

IPP 9 regulates the transfer of data outside Victoria.[35] The Transborder Data Flows principle aims to ensure that, when personal information travels, privacy protection travels with it. IPP 3 (Data Quality), IPP 4 (Data Security), IPP 2 (proper Use) and accountability all still matter when personal information is transferred outside Victoria. Victoria's privacy law will not apply to information after it is received by someone who is not subject to the *Information Privacy Act*. IPP 9 is therefore about organisations taking steps to ensure that safeguards are in place before the information leaves the protections of the *Information Privacy Act*. This can involve checking the legal protections in the other jurisdiction afforded by legislation, or taking steps to ensure that such protections are built into the contractual agreement under which information is matched (page n).[36]

Collecting organisation

NECESSITY OF COLLECTION

IPP 1.1 states that an organisation "must not collect personal information unless the information is necessary for one or more of its functions or activities". This principle is aimed at ensuring that organisations only collect information that is necessary for their purposes, and is not excessive. The issue of necessity and proportionality of collection is also relevant to consideration of whether a proposal will be consistent with the right to privacy recognised under the Charter.

As noted at page 18, "necessary" requires more than what may be administratively convenient or desired, but does not mean "essential".

It is important that the reason the organisation "needs" the information (that is, the purpose for collection) is closely tied to the organisation's function or activities. The collection should be for a specific purpose, and the type and extent of information collected should be limited to what is necessary to achieve that purpose (that is, carry out that activity or function).

Necessity will be assessed in a practical way. Does your organisation need the personal information in order to discharge its functions effectively? Consider whether anonymous information would be sufficient. IPP 8 also requires that wherever it is lawful and practicable, individuals must have the option of not identifying themselves when entering into transactions with an organisation.

Further considerations, specific to the type of public interest purpose for which you propose to disclose personal information, are set out on pages 23 to 29 of this guide.

---

[35] See *Guidelines to IPPs*, pp. 150-161.

[36] See also Privacy Victoria, *Model Terms for Transborder Data Flows of Personal Information*, June 2006.

FAIRNESS OF COLLECTION

IPP 1.2 requires that collection must be by lawful and fair means, and not in an unreasonably intrusive way.

A government organisation may not be collecting fairly if, for instance, it knowingly accepts personal information from persons whom the organisation knows are under the mistaken belief they must compulsorily provide it. Individuals may be required by law to provide certain information to obtain some benefit or entitlement, or to exercise some right or privilege–for example, to obtain a licence for a profession, or to volunteer in child-related areas of work. The legislation may set out the type of information which must be provided and, in some cases, may make it a criminal offence to provide false or misleading information. In such contexts, organisations that collect more than is necessary and compulsorily should consider carefully whether they are collecting unfairly.

SENSITIVE INFORMATION

IPP 10 will restrict the collection of personal information through data matching that falls within the definition of "sensitive information" at the start of Schedule 1 to the *Information Privacy Act*:

> "sensitive information" means information or an opinion about an individual's–
>
> (i) racial or ethnic origin; or
>
> (ii) political opinions; or
>
> (iii) membership of a political association; or
>
> (iv) religious beliefs or affiliations; or
>
> (v) philosophical beliefs; or
>
> (vi) membership of a professional or trade association; or
>
> (vii) membership of a trade union; or
>
> (viii) sexual preferences or practices; or
>
> (ix) criminal record–
>
> that is also personal information.

IPP 10 states that sensitive information must not be collected unless one of the grounds in IPPs 10.1(a)-(d) or IPP 10.2 applies.[37]

The handling of sensitive information is also subject to greater restrictions under IPP 2.1(a), which requires reasonably expected uses and disclosures to be directly related to the primary purpose for collecting the information.

INDIRECT COLLECTION

IPP 1.4 states that

> *If it is reasonable and practicable to do so, an organisation must collect personal information about an individual only from that individual.*

[37] See discussion of IPP 10 in *Guidelines to the IPPs*, pp. 162 to 170.

When collecting new information as a result of a data matching exercise, an organisation will by definition be collecting personal information other than directly from the subject individual. IPP 1.4 therefore requires a collecting organisation to justify that collecting the data directly from the individual, instead of data matching from other sources, is not reasonable or practicable.

Privacy notice issues also arise when collecting information from source agencies as part of a data matching activity. Under IPP 1.5, the collecting organisation will need to consider what "reasonable steps" it must take to ensure that the individual has been provided with a privacy notice to make them aware of the matters listed in IPP 1.3. The "reasonable steps" may range from full and direct notice (where the number of individuals affected is not large), through to general public announcements about the data matching exercise, its purposes and any further proposed disclosures. The *Information Privacy Act* makes provision for some information to be collected indirectly without giving notice, for example where such notice would pose a serious threat to the life or health of any individual, or where the collection is by a law enforcement organisation for law enforcement purposes.[38]

IDENTIFIERS

IPP 7 deals with the assignment, adoption, use and matching of unique identifiers, such as drivers' licence numbers, or unique patient identifiers. The Unique Identifiers principle aims to strike a balance between the potential privacy invasiveness of data matching, linking and profiling with the clear benefits to organisations in assigning or adopting identifiers to efficiently administer their functions.

If your organisation plans to adopt or use unique identifiers assigned by another organisation, best practice is to obtain the consent of the subject individuals.

If your organisation plans to assign a new unique identifier to the data you collect as a result of the data matching program, IPP 7.1 will be applicable:

> *An organisation must not assign unique identifiers to individuals unless the assignment of unique identifiers is necessary to enable the organisation to carry out any of its functions efficiently.*

If your organisation plans to adopt or use unique identifiers assigned by another organisation, best practice is to obtain the consent of the subject individuals.

In the absence of consent, your organisation will generally only be able to adopt unique identifiers assigned by another organisation into your own records if to do so is "necessary" to enable your organisation to carry out any of its functions efficiently.[39]

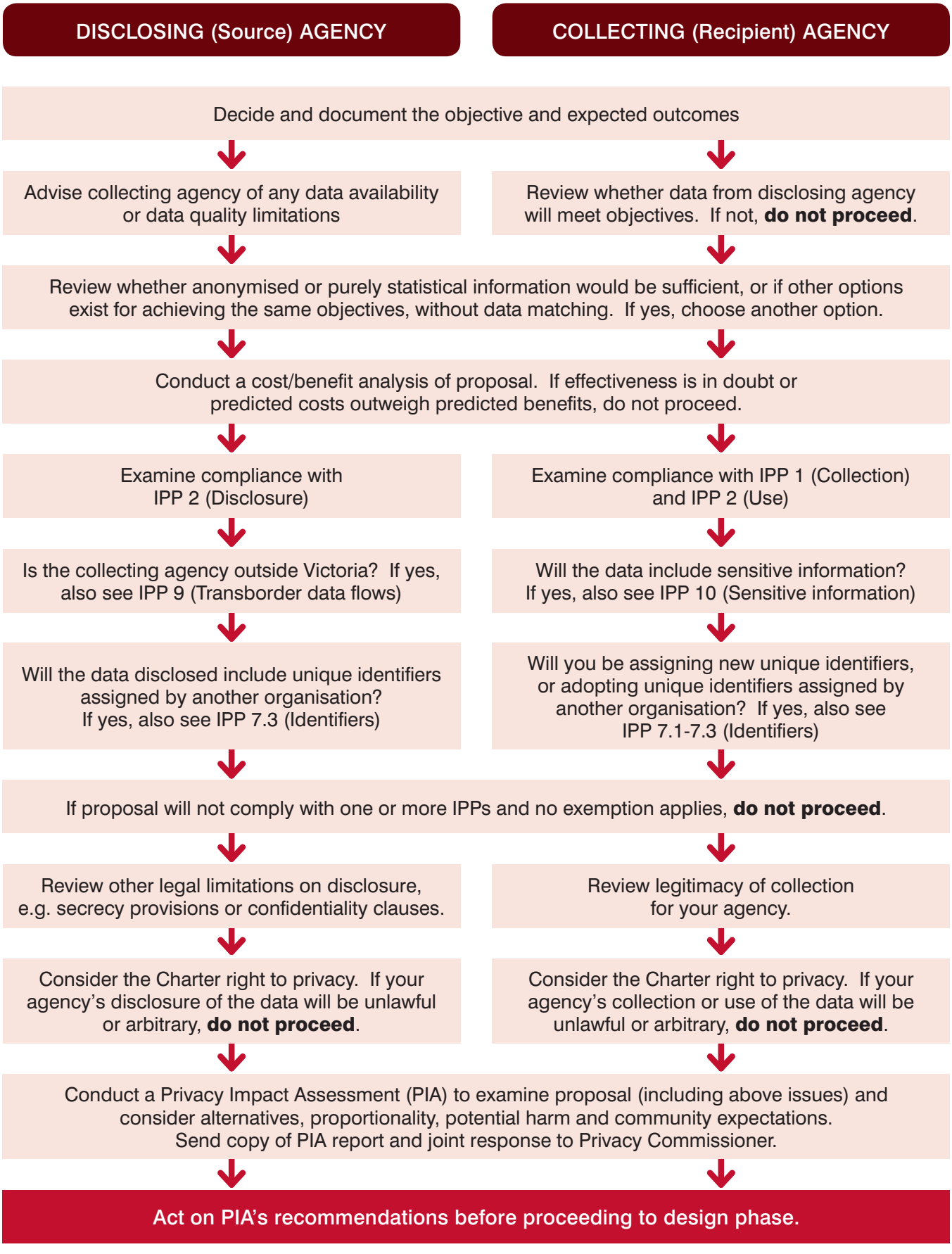
As noted at page 18, "necessary" requires more than what may be administratively convenient or desired, but does not mean "essential". However it is important that the reason your organisation "needs" the unique identifiers is closely tied to your organisation's function or activities.

---

[38] IPP 1.5 and s.13 of the *Information Privacy Act*.

[39] IPP 7.2(a). There is an additional provision which allows an outsourcing organisation to adopt the unique identifier created by a contracted service provider in the performance of its obligations to the organisation under a State contract (IPP 7.2(b)).

FIGURE 1:
Flowchart of Considerations – Concept Stage



6. Purpose

Improved efficiency or service delivery

Data matching proposals aimed at delivering improved efficiency or service delivery will usually need to meet the terms of IPP 2.1, which allows an organisation to disclose personal information if the purpose of the disclosure is related to the primary purpose of collection[40] and the individual would reasonably expect the organisation to disclose the information for that purpose.

Some motivations for data matching will be to maximise administrative and resource efficiency and to improve service delivery, often by 'cleansing' or verifying data quality. In some instances this might actually enhance and protect privacy, by improving and ensuring data quality (IPP 3), provided it is done in compliance with the IPPs.

'Data cleansing' generally involves the large-scale comparison or matching of two or more datasets, for the purposes of updating one or both of the sets. It is a relatively automatic, computerised process. However, this method carries certain privacy risks which need to be taken into account in deciding whether it would be appropriate to carry out a data matching program for the purpose of data cleansing.

The privacy risks associated with a data match may outweigh any expected benefit from improving the quality of the data held.

The potential for organisations retaining excessive information not relevant to an organisation's functions is one such risk, with its attendant risk of creating profiles of individuals drawn from the aggregation of multiple data sets. Consistent with obligations under IPP 1.1, organisations should ensure they do not collect excessive and/or irrelevant data as a result of a data cleansing exercise.

Organisations should also consider whether a data matching exercise is the best method for addressing data quality concerns. It may be more appropriate, for instance, to periodically invite individuals to update their details or to establish a mechanism for that to occur at the individual's election–such as an online service that accepts change of contact details. Providing individuals with the ability to update their data is consistent with obligations under IPP 1.4, which recognises that the best quality data is most likely to be that which is directly collected from the individual concerned. Organisations should be mindful, however, of the risks of identity crime where online change-of-details facilities are designed with a focus on convenience over security. Systems should not be designed to readily allow anyone to change any other person's details. Organisations should ensure they can authenticate the identity or authority of the person seeking to update personal information details. The authentication process needs to be as client-friendly as possible without making things easy for an identity thief.

[40] If the personal information is sensitive information, the purpose of the disclosure must be *directly* related to the primary purpose of collection.

Resultant mismatches produced from a data cleanse is another risk. Small discrepancies in a person's name or address can lead to cases of mistaken matches, and mistaken identity. It will not always be reasonable or expected to engage in data-cleansing for the purposes of updating personal information. The privacy risks associated with a data match may outweigh any expected benefit from improving the quality of the data held.

Government organisations may have inconsistent data about an individual for other reasons. Details may have changed between the different interactions with government. Names and addresses change for legitimate reasons. Marriage, divorce and, for an important minority, participation in a witness protection scheme are examples.

Organisations that match data for the purpose of data cleansing also need to consider what consequences will flow from an apparent mismatch. Whose data is more likely to be accurate? Are additional checks required before deleting or updating the record? Are there legitimate reasons for differing fields to be held about the same person? For instance, an address provided for the purpose of assessing capital gains tax may differ from the address given to an electoral commission for the purpose of voting. Legal definitions underpinning the information provided may vary. Will the 'cleansed' data lead to an action or decision that will affect the individual's rights, benefits or other interests?

These issues arose in the United States, when erroneous data supplied by a private contractor was used to cleanse the Florida voting registration rolls prior to the 2000 presidential election, resulting in the disenfranchisement of thousands of eligible voters.[41]

Reasonable steps to ensure the accuracy of personal information being used for a data cleanse, or that produced by it, might entail:

- manually double-checking the automated 'matches' for false matches due to minor discrepancies such as the wrong middle initial in a record; and
- seeking confirmation of accuracy from the individual to whom the cleansed data relates before making further use of that data.

Consistent with transparency obligations under IPPs 1.3, 1.5 and 5, organisations participating in a data cleanse should ensure they inform individuals of their practice of data cleansing– particularly where new information has been obtained during the course of the cleanse and is to be used by the organisation. Individuals should not be left wondering, for instance, how you got their new contact details.

In some cases, it will be appropriate to give all individuals whose information is involved prior notice of the matching activity to be undertaken to allow them an opportunity to seek, in advance, correction of records held. This not only facilitates data quality, but it ensures an organisation meets its obligations under IPP 1.3 (notice of collection) and IPP 1.4 (direct collection where reasonable and practicable).

---

[41] Robert E Pierre, "Botched name purge denied some the right to vote" (31 May 2001) *Washington Post*, page A01, http://www.washingtonpost.com/ac2/wp-dyn/A99749-2001May30. Also see other references made available by the Electronic Privacy Information Clearinghouse (EPIC) at http://www.epic.org/privacy/choicepoint/.

Research

Organisations collecting the results of a data matching program for the purposes of research will need to consider whether "sensitive information" (defined on page 20) is likely to be included. If so, IPP 10.2 will only allow for collection of the sensitive information if:

- it is "necessary" for research "relevant to government funded targeted welfare or educational services", and
- there is no reasonably practicable alternative to collecting the information for that purpose, and
- it is impracticable for your organisation to seek the individual's consent to the collection.

The *Information Privacy Act* facilitates research in a number of ways. Using unidentifiable data or relying on consent are alternatives to data matching which allow research to be carried out in compliance with the *Information Privacy Act*. Information can also be matched for research purposes without consent where, for example, it is authorised or required under law (IPP 2.1(f)), although in such a case the authorising legislation may itself impose obligations or restrictions on how the information is matched.

However typically, organisations disclosing personal information for the purposes of research will need to comply with IPP 2.1(c), which allows disclosures if:

- the disclosure "is necessary for research, or the compilation or analysis of statistics, in the public interest", and
- the research will not be published in an identified form, and
- it is impracticable for your organisation to seek the individual's consent, and
- your organisation reasonably believes that the recipient will not further disclose the information.[42]

An organisation seeking to rely on IPP 2.1(c) should therefore consider the following questions:

- Is the use or disclosure of identifiable information *necessary* for research or statistical work, by the organisation itself or by the proposed recipient of the information? Can the same research objectives be achieved with alternative sources of data, or data that has been de-identified or is anonymous?
- Will the research or statistical analysis/compilation result in publication of the information in a form that identifies any particular individual? If the data is to be de-identified prior to publication, how effective will that be? Consider, for instance, whether research subjects' identity can be reasonably ascertained where data is drawn from small communities.
- Does the organisation reasonably believe that the recipient of the personal information will not disclose the information? Have undertakings of confidentiality been sought? Where the disclosure is outside of Victoria, have appropriate privacy protection measures been attended to, in accordance with obligations under IPP9 (Transborder Data Flows)?
- If it is necessary to use identifiable data, can the research subjects' consent be sought? Or is it **impracticable** to seek the subjects' consent before their personal information is used or disclosed?
- Is the work in the **public** *interest*?

[42] See *Guidelines to the IPPs*, pp. 53-58.

Law enforcement

It is desirable to conduct a PIA before any large scale, systemic data matching occurs for law enforcement purposes.

IPP 2.1(e) allows disclosures of personal information if an organisation has reason to suspect that unlawful activity has been, is being or may be engaged in, and uses or discloses the personal information as a **necessary** part of its investigation of the matter or in reporting its concerns to relevant persons or authorities.

IPP 2.1(g) allows disclosures of personal information if the organisation reasonably believes that the use or disclosure is **reasonably necessary** for one or more of the following by or on behalf of a law enforcement organisation -

i) the prevention, detection, investigation, prosecution or punishment of criminal offences or breaches of a law imposing a penalty or sanction;

ii) the enforcement of laws relating to the confiscation of the proceeds of crime;

iii) the protection of the public revenue;

iv) the prevention, detection, investigation or remedying of seriously improper conduct;

v) the preparation for, or conduct of, proceedings before any court or tribunal, or implementation of the orders of a court or tribunal.

IPP 2.1(h) allows disclosures of personal information to the Australian Security Intelligence Organisation or the Australian Secret Intelligence Service, under certain conditions.

Section 13 of the *Information Privacy Act* allows for collection and disclosures for data matching by a law enforcement organisation for law enforcement purposes where non-compliance with some of the IPPs is reasonably necessary. This also applies to enforcement of laws relating to the confiscation of proceeds of crime; the conduct of current or pending proceedings in a court or tribunal; or the community policing functions of Victoria Police.

Despite this, many of the considerations discussed above and below will still be relevant. Data matching for law enforcement purposes still needs to be legitimate, necessary, lawful and proportionate in order for section 13 to apply. For this reason, where possible, it is still desirable to conduct a PIA before any large scale, systemic data matching occurs for law enforcement purposes.

In particular, disclosure of personal information in bulk may be difficult to justify as “necessary” (IPP 2.1(e)) or “reasonably necessary” (IPP 2.1(g)) if only a few individuals within the dataset are under suspicion.

Consideration should be given to the potential damage caused by incorrect matches. These could have a hugely adverse impact on the individual concerned. It should also be noted that section 13 does not grant a ‘blanket’ permission to engage in data matching without regard for the IPPs. IPPs 1.1 and 1.2, concerning collection necessity and intrusiveness, and IPPs 3 and 4, concerning data quality and data security, also still impose obligations on a law enforcement organisation involved in information matching.

For all of these reasons, before any data matching occurs for law enforcement purposes, all reasonable steps should be taken to ensure that the personal information involved is accurate, complete and up to date and that the method by which the information is matched is as secure as possible.

Protection of the public revenue

IPP 2.1(g)(iii) provides that an organisation can match personal information for a secondary purpose if the organisation reasonably believes that the use or disclosure is reasonably necessary for the protection of the public revenue by or on behalf of a law enforcement organisation.

"Public revenue" refers to regular payments to Commonwealth, State, Territory and Local Governments, such as taxes, levies, rates, application fees and charges. It is unlikely that the term covers fines enforcement, as fines are not regular payments made to a government organisation.[43] However, IPP 2.1(g)(v) may provide a basis for information matching in the fines enforcement context.

It should be noted that section 13 of the *Information Privacy Act* also permits disclosures for law enforcement agencies that believe on reasonable grounds that the disclosure is necessary for a law enforcement function or activity.

Another crucial matter to consider in this context is proportionality (e.g. how much revenue is involved?). While protection of the public revenue might amount to a justifiable and lawful reason for data matching, it may not be legitimate for an organisation to match an entire database when the protection or enforcement activity actually involves a very small percentage of the entire population. Organisations need to consider whether the protection or enforcement activity can be achieved without matching vast quantities of personal information, much of which will be irrelevant to the desired outcome.

Preparing for or responding to an emergency

IPP 2.1(d) allows the disclosure of personal information where an organisation reasonably believes that the disclosure is necessary to lessen or prevent either:

a) a **serious and imminent** threat to an **individual's** life, health, safety or welfare; or

b) a **serious** threat to **public** health, public safety, or public welfare.

The first category relates only to an imminent threat posed to an individual. This is unlikely to be relevant when considering a data matching program utilising records from multiple individuals. However the second category—something posing a serious threat to the public—may be relevant in the context of data matching.

Privacy Victoria recognises that bushfire poses a serious threat to public health, public safety or public welfare.[44]

Depending on the circumstances and the degree of notice provided beforehand, a data matching program between Local Councils and emergency services, for the purposes of safety against bushfire, may also be justified under IPP 2.1(a), which allows for secondary purposes that are related to the primary purpose of collection, and would be reasonably expected.[45]

---

[43] See *Guidelines to IPPs*, pp. 69-70.

[44] Privacy Victoria, *Bushfires and privacy*, *Info Sheet 11.02*, December 2002.

[45] Privacy Victoria, *Bushfires and privacy*, *Info Sheet 11.02*, December 2002.

The disclosing organisation should make it clear to the recipient emergency services organisation that the data is being provided for the sole purpose of ensuring public safety, not just any purpose. Emergency services may themselves breach privacy if they use or disclose the information obtained through data matching for non-safety purposes such as fundraising or membership drives.

Where a serious threat to public health or safety is involved, such as an infectious disease or large-scale evacuation, significant amounts of personal information could be at stake. Steps to ensure limited disclosure consistent with the circumstances, which may require prompt and effective action in an emergency, will need to be considered. Threats to health, safety or welfare in this context will generally require a fast and appropriate response from your organisation. Accordingly, it is advisable to have an 'emergency data matching policy' in place before it happens, and ensure all concerned personnel are aware of it.[46] That way, your organisation can quickly and confidently handle a request for personal information in an emergency situation. Such a policy may include an escalation process for dealing with such disclosures and a guide for determining who makes the disclosures, what information is likely to be released, and to whom.

Privacy Victoria has published general guidance about advance planning and training of staff in anticipation of a major event.[47] You might plan to respond to potential information requests by emergency services by starting with the following checklist, which was developed to assist Local Councils in responding to a request by authorities for personal information on their ratepayers' databases:

- What will the information be used for? The requesting emergency authority should state with reasonable precision its reasons for seeking the data. This allows the disclosing organisation to apply IPP 2 properly, and be confident about the basis for disclosure.
- Will the information meet the purpose for which it is requested? Excess disclosure can be as unhelpful as insufficient disclosure, so clarifying how much information is necessary, and tailoring the disclosure to the actual purpose, assists both the disclosing organisation and the recipient organisation. For example, if the emergency authority wants data about **residents**, then receiving **ratepayer** information may not be effective, particularly in areas with a high proportion of non-resident owner/ratepayers, as in inner Melbourne.
- Can the aim be achieved without the data proliferating? For instance, if the intended use of the information is to conduct a mail out, privacy would be better protected if the source organisation makes an appropriate arrangement with the emergency authority to mail out the information itself.

IPP 2.1(d) may also be relevant to information uses and disclosures after a disaster or accident has occurred. A proper conception of "public welfare" in this context includes offering assistance to victims, to assist the community more generally to overcome the effects of disasters and other trauma that occurs in its midst.

If new personal information is collected in order to respond to an emergency, sharing that information where necessary with other organisations involved in the disaster response may be considered disclosure for the primary purpose of collection.

---

[46] See, for example, Privacy Victoria, *Bushfires and privacy*, *Info Sheet 11.02*, December 2002.

[47] Privacy Victoria, *Local Councils and the 2006 Commonwealth Games*, *Info Sheet 06.05*, November 2005.

Data matching might, for example, assist in emergency response efforts such as locating victims and reuniting them with their family, ensuring victims receive medical attention and ensuring they have the opportunity to take advantage of various other forms of support, such as financial assistance and counselling. It is legitimate for authorities to try to reach victims to offer support. But authorities have to be aware that not everyone responds to particular authorities in the same way. Disaster victims can always decline offers of support made by or on behalf of government agencies, and their wishes for no further contact should be respected.

Personal information may also be disclosed for data matching for other 'public interest' purposes allowed under IPP 2, such as where the use is for the primary purpose of collection or for a reasonably expected related purpose. For instance, travellers who provide agencies with the contact details for their next of kin do so in order to enable contact to be made in case of emergency. During an emergency—whether that is a bombing, tsunami, train crash or other man-made or natural disaster—it would be in line with the primary reason for collecting these details for the data to be disclosed to proper authorities and used to assist in informing families in appropriate ways, and for victim identification and other relevant disaster response work.

For example, where personal information is sought after an accident or disaster occurs, IPP 2.1(a) would allow that information to be disclosed to relevant organisations involved in emergency response and recovery efforts including Commonwealth agencies such as Centrelink. Individuals caught up in a disaster would reasonably expect their information to be matched in order to locate, identify and assist them and those close to them. In this context of disaster response, privacy must be assessed with a practical eye to what might be called the exigencies of the moment. Bodies have to be identified promptly; missing or incapacitated persons need assistance quickly; and in many situations, the persons closest to the persons affected can provide authorities with important help in these tasks. In practice, this means prompt access to personal information about those affected for the proper authorities so that they can make necessary responses to an emergency.

However, regardless of whether a disclosure is justified under IPP 2.1(a) or (d), risks arise from the proliferation of data. Those risks can be minimised by taking steps to:

- keep data used or matched for the 'emergency' response quarantined from other datasets
- do not allow the data to be used for other purposes, and
- dispose of the matched data once the emergency has been responded to.

7. Considerations at the design stage

A data dictionary should be compiled to ensure a clear understanding of the correct definitions.

Defining data fields

When defining the precise data to be matched or compared, care must be taken to ensure that only data that is "necessary" is collected or disclosed. (See discussion above with respect to what 'collection necessity' and 'disclosure necessity' mean in practice) It would be useful to have both Information Technology and privacy officers involved in settling the data fields.

Before matching the information, organisations involved should ensure that they have a common definition for key information, or an agreed and reliable way to convert the information. Different organisations may record the same personal information in different ways. If this is not recognised in the design of a data matching program, the data matching exercise may lead to records becoming confused or corrupted.

For example, one organisation's definition of "income" might refer to an individual's taxable income for the last financial year, while another organisation might define "income" as current annual salary pre-tax, while a third might be referring to estimated calendar income comprising both salary and investment income streams. An algorithm will be needed to convert these data items into something that can be more readily compared–or it may simply be inappropriate to try and match this data, as a case of trying to 'compare apples with oranges'.

A data dictionary should be compiled, to ensure each participating organisation has a clear understanding of the correct definition for each data field coming from each source dataset. This will help to ensure accuracy of terms as against other datasets.

The data dictionary could also note limitations in the quality of each field of data. Metadata tags can also be useful for recording information such as the date each data field was last updated.

At this stage there should be agreement on the matching algorithm to be used, and the rules for recognising a confirmed or potential match. Whether or not statistical linkage keys (see discussion at page 8) are to be used should be settled by now.

If the decision has been made to match data that appears to be de-identified, care should be taken that it is actually **permanently** de-identified. Where it can be matched with other data held by or accessible to the collecting organisation, it might be possible for it to be re-identified, meaning that it will constitute personal information and the *Information Privacy Act* and IPPs will still apply.[48] It should also be noted that there is a significant difference between permanent de-identification before the information is matched (as it will no longer be "personal information" and subject to the *Information Privacy Act*) and a commitment by the collecting organisation to de-identify the results of the data matching after the results have been used. In the latter case, the information matched will still be "personal information" and the IPPs, including IPP 2.1, will apply.

[48] *WL v La Trobe University* [2005] VCAT 2592.

Ensuring data quality

IPP 3 requires that an organisation must take reasonable steps to make sure that the personal information it collects, uses or disclosures is accurate, complete and up to date.

It is therefore essential to check the quality of personal information before it is matched, otherwise inaccuracies, anachronisms and other problems will be spread across organisations.[49] Any plan or proposal to match information should trigger action to ensure that inaccurate records are corrected, irrelevant ones discarded and out-of-date ones updated.

Organisations should periodically check the quality of the data being matched. Privacy best practice suggests you should show your data to the subjects of that data, so that the accuracy of the information can be confirmed or disputed. Of course it is not always practical to check every record: for large datasets, a sample of records should be checked. Although this may only directly reveal deficiencies in a limited number of records, it could also indicate wider or systemic problems, which can then be addressed.

The spreading of deficient or inaccurate information across different databases and systems can cause significant problems and serious detriment for individuals. If an organisation discovers that it has disclosed flawed or inaccurate information for use in data matching, it should not only correct its own records, but also ensure that the information is corrected by the recipients of that information.

There also needs to be procedures in place for dealing with situations where there are disagreements between participating organisations about the accuracy, completeness or freshness of records. In some cases, the best thing to do may be to ask the individual concerned whether the results of the data matching are correct.

It is the primary responsibility of the disclosing organisation to ensure that the data will continue to be adequately secured once other organisations have access to it.

Ensuring data security

Under IPP 4.1, organisations are obliged to take reasonable steps to protect information they hold from misuse, loss, unauthorised access, unauthorised modification or unauthorised disclosure.[50] While it is not directly specified in IPP 4, the key consideration should be to make sure that security is adequate in relation to the potential damage that a security breach could cause.[51] More sensitive or confidential information therefore needs a higher level of security.

It is the primary responsibility of the disclosing organisation to ensure that the data will continue to be adequately secured once other organisations have access to it. There should be arrangements in place that set out the security that will apply to the information once it has been matched.

Data security considerations at the design stage should include:

---

[49] UK Information Commissioner's Office, *Framework code of practice for matching personal information,* October 2007, pp. 11-12.

[50] See *Guidelines to IPPs*, pp. 90-106.

[51] UK Information Commissioner's Office, *Framework code of practice for matching personal information,* October 2007, p. 15.

- Will all the organisations involved adopt a common security standard?
- Will media be involved (e.g. CDs, DVDs) in transit?[52]
- Will the matching be done using online mechanisms?
- Is the access for online data matching secure?
- Is authentication or encryption used during transmission?[53]
- Are there policies and/or technical controls in place for the use of portable storage devices and other technology used for data matching?[54]
- What background vetting will be done of staff?
- What training will be provided to staff about their privacy obligations?

It is particularly important to ensure that where private sector organisations are involved in the data matching activities of government organisations, the privacy protection is clear and enforceable.

Under IPP 4.2, organisations are also obliged to take reasonable steps to destroy or permanently de-identify personal information if it is no longer needed for any purpose (see "Data Retention and Data Disposal" at the page 39). Organisations involved in a data matching program should, at the design stage, make an agreement about what should happen once the need to match the information has passed. This agreement should be incorporated into the Data Matching Protocol (see page 33).

Accountability

Before a data matching project commences, it should be determined who will be responsible for dealing with compliance issues that arise. While the disclosing (source) organisation has the primary responsibility under IPP 4 to ensure that the data is handled properly, even after it leaves its custody and control, it is also important to ensure that every participating organisation is subject to enforceable privacy standards. It is particularly important to ensure that where private sector organisations are involved in the data matching activities of government organisations, which may mean access to large amounts of compulsorily collected personal information, the privacy protection is clear and enforceable.

Only the Victorian public sector organisations involved, or their "contracted service providers", will have responsibility under the IPPs in the *Information Privacy Act*.

---

[52] Ensure secure handling practices are used, as media itself cannot be suitably protected

[53] The New Zealand Privacy Commissioner now requires encryption for physical transfers of personal information using digital media, when approving "information matching" exercises under the *Privacy Act 1993 (NZ)*. See http://www.privacy.org.nz/assets/Files/IM-Bulletins-and-related-material/IM-Bulletin-2008-06.pdf, last accessed 21 July 2008.

[54] See Privacy Victoria, *Use of Portable Storage Devices*, January 2009. In November 2007, Her Majesty's Revenue & Customs in the United Kingdom lost two unencrypted disks with personal data on 25 million British citizens, including dates of birth, addresses, bank account information and national insurance numbers. In January 2008, a Ministry of Defence laptop was stolen from the car of a military recruitment officer, containing information about approximately 600 000 people, most of whom were prospective recruits. The database stored on the laptop was unencrypted. See http://www.ico.gov.uk/upload/documents/pressreleases/2008/data_security_220408_2.pdf, last accessed 21 July 2008.

The *Privacy Act* regulates some private sector organisations under the National Privacy Principles (NPPs), but there are currently significant gaps in coverage, such as for business with less than a $3M pa turnover. Be aware of other exemptions. For example, even for large businesses, services provided to another State government under State contracts (i.e. those activities conducted as "contracted service providers") are exempt from the NPPs.

Privacy Victoria's 'snapshot' audit of data matching activities in the Victorian public sector found that among the participating organisations there was considerable movement of data between organisations and contracted service providers.[55] It is now common for public sector agencies to outsource functions, including IT and other information processing activities, to contracted service providers.

If a participant in a data matching program is a "contracted service provider" as defined in sections 9 and 17 of the *Information Privacy Act*,[56] particular considerations apply. Firstly, the relationship and contract under which services are to be provided will need to be carefully examined or drafted to ensure that the body is in fact a "contracted service provider". Then care will need to be taken to properly utilise section 17(2) of the *Information Privacy Act*, to provide that the contracted service provider is bound by the IPPs for any act or practice conducted under the State contract, in the same way that the outsourcing organisation would have been bound.

If a private sector matching organisation or collecting organisation does not meet the definition of "contracted service provider", the disclosing organisation needs to take additional steps to properly protect the information being matched.

The same concern applies where a recipient organisation is from another State or Territory government. If the data is to be disclosed outside Victoria, to a jurisdiction without enforceable privacy standards equivalent to the IPPs, IPP 9(f) obliges the disclosing organisation to take reasonable steps to ensure that the data will not be held, used or disclosed by the recipient inconsistently with the IPPs.

In such situations, the disclosing organisation needs to take steps to ensure that privacy protections and accountability/indemnity arrangements are built into the contractual agreement underpinning the Data Matching Protocol.[57]

Data Matching Protocol

Privacy Victoria's audit of data matching activities in the Victorian public sector attempted to find out how organisations protected data from misuse when it was provided to external organisations for data matching purposes. A number of organisations had written agreements with the receiving organisation.[58]

---

55 Privacy Victoria, *Victorian Public Sector Data Matching Audit*, February 2005.

56 See also Privacy Victoria, Information Sheet 1.06, *Who's Covered by the Information Privacy Act?*, 20 April 2006.

57 See Privacy Victoria, *Model Terms for Transborder Data Flows of Personal Information*, June 2006.

58 Privacy Victoria, *Victorian Public Sector Data Matching Audit*, February 2005.

Mostly, the written agreements for data matching took the form of a Memorandum of Understanding (MOU). An MOU is not enforceable. It is not a contract in law. An MOU may be useful in reducing to writing precisely what the various parties to the data match expect to do and receive. From this perspective, MOUs may assist in compliance with IPPs 1.3 (Notice), 5 (Openness) and 2 (Use and Disclosure—in the sense that purposes are clear) if made public. However MOUs are generally not sufficient in themselves to ensure compliance with IPPs 4 (Data Security) or 9 (Transborder Data Flows).

If any of the matching or recipient organisations are not Victorian public sector organisations regulated by the *Information Privacy Act*, we recommend use of a contract, rather than an MOU.

For this reason, if any of the matching or recipient organisations are *not* Victorian public sector organisations regulated by the *Information Privacy Act*, we recommend use of a contract, rather than an MOU.

Regardless of whether an MOU or contractual format is used, a Data Matching Protocol should be developed to document the agreements between the various parties. A Data Matching Protocol should help define the roles and obligations of the various parties, and spell out the 'nuts and bolts' of how the data matching exercise will work.

The Data Matching Protocol should document such matters as:

- the 'public interest' objective of the program;
- each organisation involved;
- for each disclosing organisation, which permitted disclosure under IPP 2 applies;
- for each organisation, what data fields they will be disclosing, matching and/or collecting;
- the data dictionary;
- notes on any data quality limitations;
- the matching algorithm/s to be used, and the rules for recognising a confirmed or potential match;
- who will conduct the match and where, and the destination for the match results;
- the purposes for which the results of the data matching can and cannot be used by the recipient organisation/s, and any prohibition on further disclosure;
- what additional steps will be taken to verify the accuracy of the data before any adverse action is considered;
- what adverse action may be taken against affected individuals;
- privacy notice requirements for each organisation involved;
- data security requirements for each organisation involved;
- data retention schedules and data disposal requirements for each organisation involved;
- the designated officer/s to receive and handle complaints and enquiries, including access requests and correction requests;
- if any of the matching or recipient organisations are *not* Victorian public sector organisations regulated by the *Information Privacy Act*—the contractual arrangements to assign accountability for privacy protection;
- how the agreement will be enforced if one of the participating organisations breaches the MOU / contract, including applicable sanctions; and
- arrangements for evaluation, review, and audit of the data matching program, including reporting and the conduct of PIAs.

The Data Matching Protocol should be a public document, although it may be appropriate to redact (mask) some of the more detailed information relating to the data fields and algorithms prior to publication, particularly if the purpose of the data matching program relates to law enforcement and data security purposes.

In 1998 the Office of the Federal Privacy Commissioner published a suggested format for a Data Matching Protocol, with more detail than we have outlined here. A copy of that guidance has been reproduced with permission in the Appendix to this guide, with modifications appropriate to the Victorian legislative environment.

Transparency and giving notice

Transparency is an important part of privacy regulation. Notice under IPP 1 is important for transparency as is the openness required by IPP 5. While there can be various approaches to address the issue of giving individuals notice of a data matching program, a notice should achieve all of the following:

- the fact, plainly stated, that the individual's personal information would be matched with other personal information about the individual held in another record and (perhaps) collected from the individual by another organisation;
- the purposes of the data matching activity;
- to whom the organisation usually discloses information for data matching; and
- any law that requires the particular information to be collected or disclosed for the purpose of data matching.

This type of privacy notice should be provided to individuals at the time their personal information is collected (if it is known at the time of the collection that their information will or might be used or disclosed for data matching), or later, by writing to the affected individuals directly or by publishing notices in newsletters or other publications likely to reach the intended audience.

Where consent is being relied upon under IPP 2.1(b), forms for documenting consent to data match should also be included as part of the privacy notice. It will usually be more practicable to obtain consent at the time of collection, but care should be taken to ensure that consent remains specific, informed, voluntary and current. Consent will not be genuine unless a refusal to provide consent means the data matching will not go ahead in relation to that individual's data.

IPP 5 requires an organisation to set out in a document clearly expressed policies on its management of personal information. The organisation must make the document available to anyone who asks for it (IPP 5.1). On request by a person, an organisation must take reasonable steps to let the person know, generally, what sort of personal information it holds, for what purposes and how it collects, holds, uses and discloses that information (IPP 5.2).[59]

IPP 5 notices, or 'privacy policies' as they are sometimes known, are intended to inform the general public how personal information will be managed, including how it will be matched and what it will be used for. An organisation engaged in data matching needs to decide whether a single document is enough to inform the public of all the ways in which personal information is managed.

---

[59] *Guidelines to IPPs*, pp. 112-115;

Privacy best practice would suggest that there should be a separate policy document for each particular data matching initiative, that would allow much more detailed and specific information to be provided. A publicly available copy of the Data Matching Protocol can serve this purpose well.

There are cases where it is legitimate to match information without a person's knowledge. This can be the case in, for example, the child protection context or to safeguard public safety in an emergency. In many law enforcement contexts, it is not possible to provide specific information about data matching, as doing so may prejudice an investigation. However, even in these contexts, organisations should be open with the public about the sorts of circumstances in which information may be matched without knowledge or consent.

Privacy best practice would suggest that there should be a separate policy document for each particular data matching initiative, that would allow much more detailed and specific information to be provided.

It is generally advisable to publish information on the data matching activities to the fullest extent possible. Make policies and protocols accessible, as well as the complaints procedure and the person to whom complaints should be addressed. Transparency will generate trust and confidence, and enhance data quality over time.

Pilot program

Once all the conceptual and design issues have been resolved and documented, a pilot program should be considered before the intended data matching program is commenced.

The purpose of a pilot program is to:

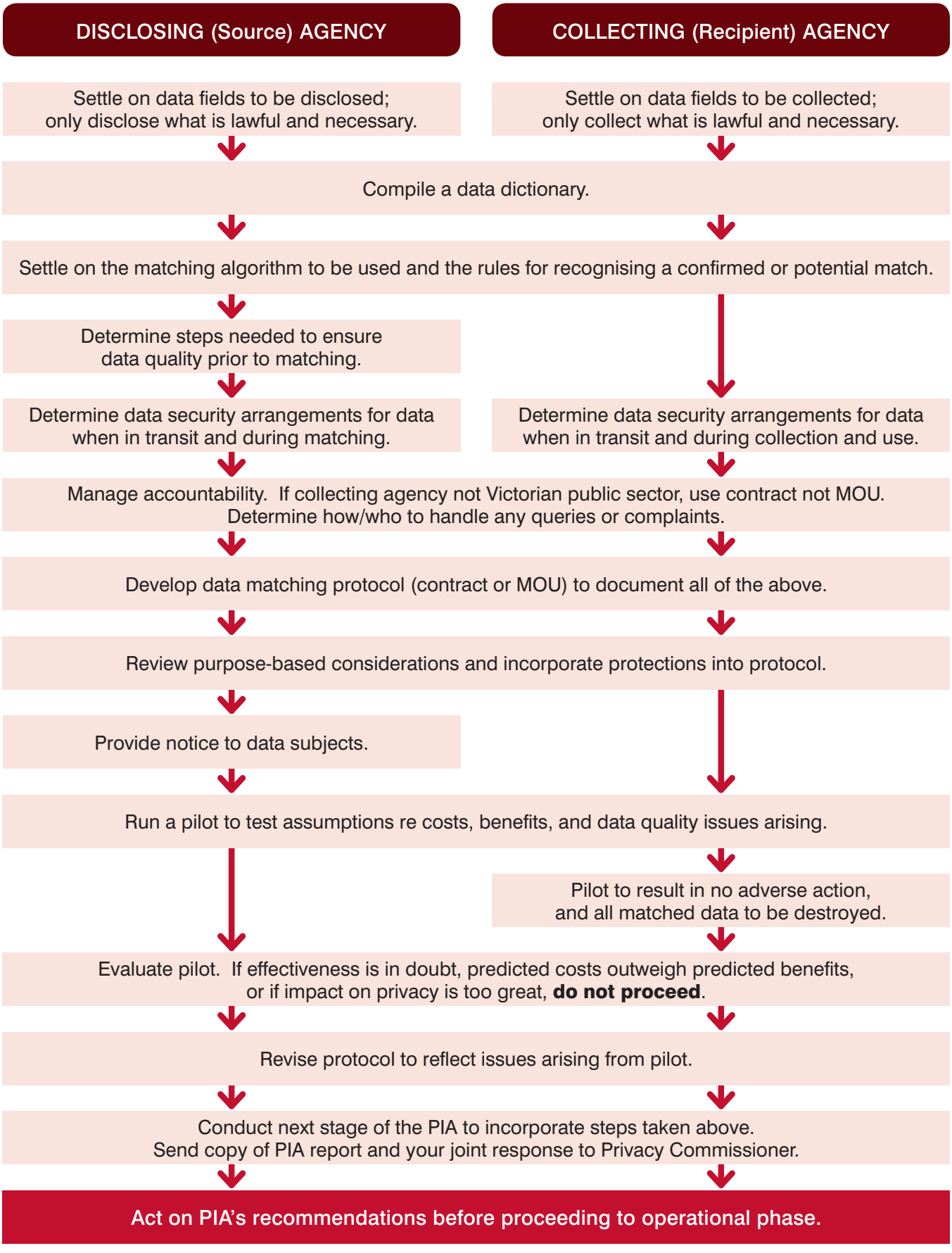
- test that the operational arrangements (such as data definitions and the matching algorithms) work in practice;
- test assumptions about the costs and benefits of the proposed program;
- identify any issues or concerns that require further clarification or resolution; and
- test the adequacy of data quality and data security arrangements.

A pilot program should involve only a small sample of the datasets intended to be included in the 'full' data matching program. The data used in a pilot program should remain quarantined from organisation's other data holdings, and should not be used for any other purpose. In particular, data used in a pilot program should not result in any administrative action in relation to subject individuals. However the pilot program should be subject to the same data security and accountability protections as if it were the 'full' data matching program.

The pilot program should be subject to a comprehensive evaluation. If the effectiveness of the program in achieving the 'public interest' purpose is in doubt, or if predicted costs will outweigh predicted benefits, or if the impact on privacy will be too great, the data matching program should not proceed any further.

The data used in a pilot program should ideally be destroyed as soon as the pilot has been evaluated, although organisations will need to consult with the Public Record Office Victoria (PROV) before destroying any records. If the program is intended to proceed, the Data Matching Protocol should first be revised to reflect any issues arising from the pilot. The PIA should also be revised to reflect issues arising from the pilot, or other operational issues settled since the previous PIA was conducted.

FIGURE 2:
Flowchart of Considerations – Design Stage



8. Considerations at the operational stage

Conducting the data match

The data being matched should first be extracted from the source datasets, and kept physically separate from operational records until the matching and checking processes are complete. Unverified information must not be added to an agency's records until it is confirmed that the data is accurate, necessary and relevant.

The data matching algorithm will establish what constitutes a successful match. For example, the algorithm may consider a match to be cases where the full name, date of birth and address are all the same, or it may allow for 'likely' matches if the surname and date of birth are the same, even if the first name differs slightly. In some cases, it might be the *absence* of a person from one of the datasets that is of interest.

The process will normally produce pairs of records which are judged **likely** to relate to the same person (or the absence of a pair where there should be one), but these matches cannot be said to be **certain**. The algorithm to be used requires careful thought and practical trialling before implementation; too 'tight' an algorithm will miss many matches of records which are actually about the same individual, and too 'loose' an algorithm will pair an unacceptably high proportion of records which are really about different individuals.

Once the data matching exercise has been run, information that did not produce any pairs of records (matches) should be destroyed immediately. The matching process will produce a list of 'raw matches', which should be put through confirmation procedures. This may involve a manual check of the original records held by the source organisation/s. The confirmation procedures may reveal some mismatches, which should then also be destroyed.

If the resultant checked matches are to be used as the basis for taking adverse action against individuals (such as reducing or cancelling a benefit, or imposing a penalty), they should be subject to further verification of results. The confirmation procedures outlined above do not necessarily 'prove' that the two records are about the same person.

It is therefore not advisable to act on the basis of an apparent discrepancy produced by a match, even with some in-house checking completed. In fairness, the information should be verified by being shown to the individual concerned before action is taken. This allows an opportunity for the data to be disputed. People should not be 'presumed guilty' solely on the basis of inferences drawn from a data matching process. Notice is an especially important safeguard where the matching process might have wrongly associated records relating to different individuals.

In fairness, the information should be verified by being shown to the individual concerned before action is taken.

Ideally, written notice should be provided to the individual, giving details of the discrepancy and the proposed adverse action, and allowing the individual a suitable time period from receipt of the notice to show reason why such action should not be taken.[60] Nonetheless, if the data matching results are to be used as the basis for taking adverse action against individuals, they should be acted upon in a timely fashion. The information should not be allowed to become out of date, as this could prejudice the individuals concerned.

For checked matches, once the notice period has expired and no successful challenge has been launched by the individual concerned, the recipient organisation can take appropriate administrative action, and incorporate the new data into agency records if necessary.

If the resultant checked matches are *not* to be used as the basis for taking adverse action against individuals, once they have served the purpose of the data matching program (such as research or law enforcement), the data should be destroyed, subject to any data retention requirements outlined below.

Data retention and data disposal

Under IPP 4.2, organisations are obliged to take reasonable steps to destroy or permanently de-identify personal information if it is no longer needed for any purpose.[61]

IPP 4.2 does not authorise retention of information 'just in case' it is needed for some future use by the organisation or by a third party. Transparency and fair collection help maintain public confidence that personal information is to be used in accordance with assurances given at the time of collection. The potential for misuse or secondary uses of data or 'function creep' is therefore curtailed where the period of data retention is strictly controlled.

Subject to the *Public Records Act 1973 (Vic)*,[62] personal information should therefore not be kept for longer than is necessary. Data disposal can be by destruction or permanent de-identification.

IPP 4.2 will not authorise the disposal of records that must be preserved under the *Public Records Act*. If your organisation is bound by the *Public Records Act*, you should ensure you have the necessary authority from PROV before destroying or altering (through de-identification) any public records. If a relevant Records Authority from PROV does not apply to a particular record or class of record in your custody, then you should request an appraisal from PROV to determine the appropriate disposal action.

In some cases, a Records Authority issued by PROV will set a minimum period for retention. In other cases, records can be destroyed without further authorisation from PROV under the "normal administrative practice" (NAP) principle, which applies to records of a facilitative ephemeral nature (such as rough notes for preparing correspondence and other records). Where these records contain personal information, you should consider if there is any legitimate purpose for retaining the records beyond the minimum retention period or despite the NAP principle. If not, then IPP 4.2 applies and you should take reasonable steps to destroy or de-identify the information.

---

[60] The federal Privacy Commissioner's guidelines mandate a time period of at least 14 days; see Office of the Privacy Commissioner, *The use of data matching in Commonwealth administration – Guidelines*, February 1998, p.12, available at http://www.privacy.gov.au/act/datamatching/index.html.

[61] See *Guidelines to IPPs*, pp. 106-111.

[62] See http://www.prov.vic.gov.au.

The actual matching process should be done in a quarantined environment—meaning that the data to be used in the matching process should be copies of original data sets, not the original data sets themselves. As soon as the data matching has occurred, only the 'matched' records should be retained as a new dataset; the rest of the source data (i.e. unrelated records) should be disposed of immediately.

The actual matching process should be done in a quarantined environment.

The various organisations involved need an agreement about what should happen with the new dataset created from the matching exercise. Different organisations may need to set their own data retention periods for the information. However, if matched information should be deleted altogether, for example because it is no longer relevant for the program's purposes, then all of the organisations holding copies of the information should delete it. If the information has a statutory retention period that has been exceeded, the source organisation must ensure that any recipient organisation also deletes it, or permanently de-identifies the data.

In temporary data matching situations such as emergencies, the best course of action will be for all recipient organisations to destroy their copies of the results of the data matching exercise as soon as the particular 'public interest' purpose has been met. Data created or collected for the purposes of responding to an emergency should not be incorporated into 'regular' records.

Data disposal must be thorough. Care should be taken to ensure that any de-identification process is truly permanent and not reversible. It can also be all too easy to overlook the retention of paper records in archives or filing cabinets.

"Destruction" may be guaranteed only if all relevant hardware and corporate backup media are destroyed—neither of which is likely. "De-identification" is very difficult to achieve, as all data fragments may not be locatable in the corporate information architecture. Accordingly, and as stated above, it is preferable to keep the results of the data matching quarantined from the regular network, with its own system of backups. Those specific backup media can then be destroyed along with the data at the appropriate time.

Handling queries or complaints

There should be clear protocols in place for handling queries or complaints from individuals. Organisations should consider whether their policies and procedures for handling general privacy complaints are sufficient, or whether specific protocols for the data matching exercise should be put in place.

It may be appropriate that complaints concerning acts and practices of one of the parties to the data matching agreement be handled by that party. Complaints about practices of multiple parties or the exercise as a whole may need to be directed to a designated body/steering committee. Consider including in the Data Matching Protocol a provision that the source organisation must be notified of any complaints.

There should be clear protocols in place for handling queries or complaints from individuals.

It is good practice for organisations to have systems in place for dealing with enquiries about information matching in a timely and helpful manner. An analysis of questions (and complaints) can help an organisation respond to public attitudes to the data matching being carried out, and to make any necessary improvements.

Evaluation, review and audit

While conducting or commissioning a PIA prior to undertaking a data matching program is an essential component of sound risk and project management, so too is evaluation and review during and/or after the data matching. Evaluation and review should be used to gauge whether the underlying principles of legitimacy, necessity, lawfulness, proportionality, effectiveness and transparency are being met, and whether the program is in compliance with the IPPs and the Charter right to privacy.

One way to effect this is to audit the data matching program, externally or internally. Privacy Victoria has published an *Audit Manual*,[63] which can be used by organisations to assist in this process.

Organisations will need to consider how frequently the data matching program should be evaluated or reviewed. This will vary, depending on the nature of the data matching exercise and the information being matched. Large scale, systemic data matching exercises will generally require frequent, periodic reviews, as part of an appropriate risk management framework.[64] A fresh PIA is recommended if the evaluation recommends that the data matching program be repeated or become routine, particularly where multiple jurisdictions are involved.

An essential part of evaluation and review is the ability to make necessary changes in light of the evaluation. If any of the underlying principles are found to no longer be valid, or there are identified IPP breaches or unjustified interferences with privacy, the data matching may need to be completely suspended and steps taken to deal with any harm caused.[65]

Finally, to help maintain transparency and accountability for a data matching program, best practice suggests that you send a copy of the evaluation report to the Privacy Commissioner, and include a summary in your Annual Report.

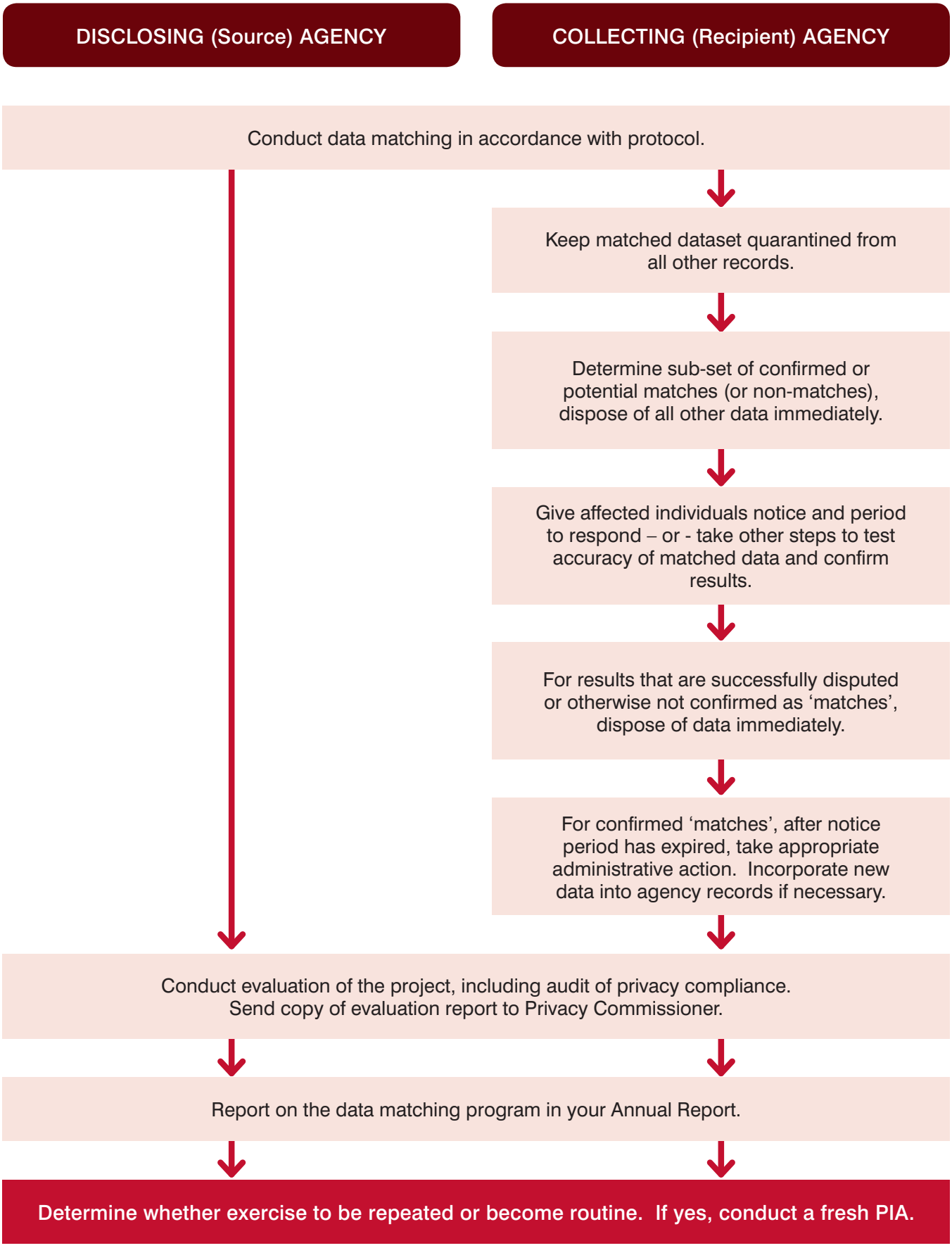
---

63 Privacy Victoria, *Privacy Audit Manual*, November 2007, Edition.02.

64 The federal Privacy Commissioner's guidelines require evaluation of data matching programs at least every three years; see Office of the Privacy Commissioner, *The use of data matching in Commonwealth administration – Guidelines*, February 1998, p.14, available at http://www.privacy.gov.au/act/datamatching/index.html.

65 See Privacy Victoria, *Responding to Privacy Breaches, Guide Edition 1,* May 2008.

FIGURE 3:
Flowchart of Considerations – Operational Stage



APPENDIX:
Checklist of matters to include in a Data Matching Protocol

In 1998 the Office of the Federal Privacy Commissioner published a suggested format for a Data Matching Protocol (program protocol).[66] A copy of that guidance has been reproduced with permission here, with modifications appropriate to the Victorian legislative environment.

This checklist is intended only as a suggestion, and should be read in conjunction with the section on Data Matching Protocols earlier in this guide.

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 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 detail 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-Victorian public sector sources; and
- all agencies or other organisations that have access to the results of the program.

---

[66] Office of the Privacy Commissioner, *The use of data matching in Commonwealth administration – Guidelines*, February 1998, available at http://www.privacy.gov.au/act/datamatching/index.html; see Appendix A, "Content of data-matching program protocols", pp. 14-17.

(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 (e.g. driver's licence number, name and date of birth), what criteria are used to identify a 'match' (e.g. individuals on both files, individuals on one but not the other) and the fields included in each output file.

(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 template 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;
- 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 IPP 1.3 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 *Information Privacy Act*, Charter, and any other relevant legislation.

To be lawful, any disclosure of personal information by an agency must fall within one of the exceptions to IPP 2 in the *Information 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 permitted disclosures in IPP 2.

The protocol must specify which permitted disclosures in IPP 2 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)

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. Be wary of reliance upon intangible benefits.

(xiv) Approval

This should be at a sufficiently senior level.

Victoria's Information Privacy Principles (IPPs) Summary

1. Collection

   Collect only personal information that is necessary for performance of functions. Advise individuals that they can gain access to their personal information.

2. Use and disclosure

   Use and disclose personal information only for the primary purpose for which it was collected or a secondary purpose the person would reasonably expect. Uses for secondary purposes should have the consent of the person.

3. Data Quality

   Make sure personal information is accurate, complete and up to date.

4. Data Security

   Take reasonable steps to protect personal information from misuse, unauthorised access, modification or disclosure.

5. Openness

   Document clearly expressed policies on management of personal information and provide the policies to anyone who asks.

6. Access and correction

   Individuals have a right to seek access to their personal information and seek corrections. Access and correction will be handled mostly under the Victorian *Freedom of Information Act*.

7. Unique identifiers

   A unique identifier is usually a number assigned to an individual in order to identify the person for the purposes of an organisation's operations. Tax File Numbers and Driver's Licence Numbers are examples. Unique identifiers can facilitate data matching. Data matching can diminish privacy. IPP 7 limits the adoption and sharing of unique identifiers.

8. Anonymity

   Give individuals the option of not identifying themselves when entering transactions with organisations, if this would be lawful and feasible.

9. Transborder data flows

   Basically, if your personal information travels, privacy protection should travel with it. Transfer of personal information outside Victoria is restricted. Personal information may be transferred only if the recipient protects privacy under standards similar to Victoria's IPPs.

10. Sensitive information

    The law restricts collection of sensitive information like an individuals racial or ethnic origin, political views, religious beliefs, sexual preferences, membership of groups or criminal record.

The full text of the Information Privacy Principles forms schedule 1 of the *Information Privacy Act 2000 (Vic).*

To determine legal rights and responsibilities, use the full version, not this summary.

The Information Privacy Principles are simply…

the right information,
to the right people,
for the right reason,
in the right way,
at the right time.

Office of the
Victorian Privacy
Commissioner

GPO Box 5057
Melbourne Victoria 3001
Australia
DX 210643 Melbourne

Level 11
10-16 Queen Street
Melbourne Victoria 3000
Australia

Local Call 1300 666 444
Local Fax 1300 666 445

www.privacy.vic.gov.au
enquiries@privacy.vic.gov.au