Public Record Office Victoria
Standards and Policy

Recordkeeping Policy

PUBLIC RECORD
OFFICE VICTORIA

Use of Back-up Technology to Archive

Issues Paper

PUBLIC RECORD
OFFICE VICTORIA

Acronyms

The following acronyms are used throughout this document.

PROV	Public Record Office Victoria
RDA	Retention and Disposal Authority

Table of Contents

Copyright Statement

Disclaimer

General

The State of Victoria gives no warranty that the information in this version is correct or complete, error free or contains no omissions. The State of Victoria shall not be liable for any loss howsoever caused whether due to negligence or otherwise arising from the use of this issues paper. This issues paper should not constitute, and should not be read as, a competent legal opinion. Agencies are advised to seek independent legal advice if appropriate.

Records Management Standards Application

The recordkeeping standards issued by PROV apply to all records in all formats, media or systems (including business systems). Agencies are advised to conduct an independent assessment to determine what other records management requirements apply.

Executive Summary

This issues paper proposes the following position on the use of back-up technology to archive data within Victorian public offices.

Archiving is the process of ensuring that data is kept accessible for future use, even though day-to-day use of the data has ceased. Back-up technology is used to recover data over the short term when it has been corrupted or destroyed by hardware failure, software failure, operator error, or malicious action. Back-up is effective at short term recovery, but archiving is about long term access. This conflict in time frames means that there is a significant risk when using back-up technology to archive that the data will not be able to be recovered. The risk is caused by the need to have both the back-up software and the original application in order to extract meaning from the backed-up data. Consequently it is proposed that the use of back-up technology to archive data will not meet PROV's standards.

It is also proposed that PROV will not, in general, recommend the use of individual media (such as tape) to archive data. Individual media should only be used when the quantity of data makes other storage uneconomic. If individual media is used to archive data, an agency should institute a management regime to ensure that media is not lost, that the condition of the media is tracked, and that the information is copied off media before the media deteriorates or the technology becomes obsolete.

Agencies that do not keep data accessible for the required periods may be infringing citizen's entitlements, not supplying the quality of service they are required to, be at risk in legal proceedings, be unable to demonstrate appropriate governance and accountability, and be affecting the right of the Victorian public to understand their history. The retention period depends on the purpose of the data, and can range from immediate destruction after use to being kept permanently. A significant amount of data needs to be kept for a decade or more.

The period for which data must be kept accessible is determined by agencies, and approved by the Keeper of Public Records. These determinations are published in Retention and Disposal Authorities (RDAs) available from the PROV website.

The consultation phase will conclude on **31 December 2012**. The comments received will inform an official advice from PROV regarding the use of backup technology for archiving.

Please send comments to standards@prov.vic.gov.au

1. Introduction

1.1. Purpose

This issues paper proposes a position on the use of back-up technology to archive data within Victorian public offices.

PROV invites input from agencies and other stakeholders to inform the development of the PROV *Use of Backup Technology to Archive Data* policy.

This policy will be binding on Victorian government agencies as it reflects principles in the PROV Disposal and Storage Standards[1]. Under the Public Records Act (1973) the Keeper has the power to set standards for the efficient management of public records. These standards apply to all records created by the Victorian Government.

1.2. Scope

The scope of this issues paper is to clarify the risks and responsibilities of an agency using backup technology to 'archive' data.

In this issues paper the term 'data' is considered to be synonymous with record. The *Public Records Act (1973)* defines a record to be a document as defined in the *Evidence Act (2008)*, and a public record as any record made or received by a officer in the course of their duties. Almost any data, or electronic document held within an agency would be covered by these definitions.

1.3. How to respond

This issues paper invites comments from all interested parties.

The consultation phase will conclude on **31 December 2012.**

The comments received will inform an official policy on the *Use of Backup Technology to Archive Data*.

Please send all comments or questions to standards@prov.vic.gov.au.

[1] These Standards are available at http://prov.vic.gov.au/government/standards-and-policy/disposal and http://prov.vic.gov.au/government/standards-and-policy/storage

2. The distinction between archiving and back-up

Data archiving is keeping data accessible for as long as it is required. Accessible means that the information contained within the data can be extracted and used. Data may need to be accessible for periods ranging from days, through decades, to indefinitely.

A back-up is a copy of all or part of the data for the purposes of recovering the data in the event of a disaster, hardware or software failure, operational error, or malicious activity. Implemented correctly, it is an effective strategy to guard against these risks, however, the focus is on short term recovery of data.

2.1. Questions:

1. Do you have any comments on the distinction we are drawing between archiving and backing-up data?

3. Issues with using back-up technology to archive data

Back-up technology is designed to recover data that's a couple of weeks old but which has been lost or corrupted due to hardware failure, software failure, operator error, or malicious activity. It is a short term risk management strategy. Data stored in a backup system beyond the life of the creating system, or the system used to back up the creating system, is vulnerable.

Essentially the problem is that the backed-up data needs to be interpreted by two layers of applications in order to recover meaning (the required information). The backed-up data first needs to be interpreted by the back-up software in order to restore the data as it was when the back-up was created. Then the application that used the data must be run to interpret the data and provide access to it. The Public Record Office Victoria (PROV) considers that it will be unlikely that both of these applications will be available after a reasonable period to recover the 'archived' data.

3.1. Lack of the back-up software

In order to restore the backed-up data, it is normally necessary to have access to the back-up software (and often the specific version of the software) that originally created the back-up image.

The need to have access to the original back-up software is because the data formats used by back-up software are normally proprietary and complex. This, in turn, is because the competitive advantage of commercial back-up products is, first, that back-ups can be completed in a short window (to minimise application downtime), and, second, that the resulting data is small in size (to minimise storage requirements).

That back-up images are complex and proprietary is normally not a concern when using back-up technology to protect against disasters, errors and failures, as the period between creating the back-up and restoring it is normally short (say less than a year), and so the back-up software is normally available.

However, when archiving data the gap between creation of the back-up and restoration may be decades. Even if an agency continued to license the back-up software for this period, which would be expensive, it is likely that the underlying back-up formats would evolve and the latest version would be unable to restore

old images. Over a long period of time, the vendor is likely to cease supporting the product, or simply go out of business.

The back-up software could, of course, be archived itself (although this would normally require ongoing licensing). But changes in the computers or operating systems mean that it is likely that the archived back-up software could not be run on current computers. This is particularly true where the back-up software operates below the level of the operating system (e.g. directly copying disc blocks).

A very small number of back-up technologies do not suffer from this issue. For example, the Posix TAR program (and its successors) has a standardised data format. However, these formats are not widely used today in government IT environments as they do not give short back-up windows and high compression.

3.2. Lack of the application software

Even if the back-up software is available and can restore the data image onto disc, this normally does not provide access to the archived data. Instead, it provides access to an image of the data as held by the original application at the time the back-up was made. In order for the data to be accessible an application needs to be available to interpret the restored data and provide access to the required information.

The need for an application to interpret the stored data to present the information is normal for accessing digital records – to access the information in a Word file requires either a version of Microsoft Word, or an equivalent program that reads Word files and accurately displays them.

Accessing stored data is more complex than accessing a Word file. The organisation of stored data from a business application has a generic component (how the underlying database application stores tables and represents data), and a component specific to that business application (the meaning of the tables and columns within the database, and any programmatic interpretation of the data). This means that it is not just sufficient to have access to the generic database software (as this will not give the business specific information), or the business application design (as this will not give the translation of the logical database design to structures on disk). Both the software and the design are necessary to access the data.

Again, this is not a concern when using back-up technology for its intended purpose, to recover from short term data failures, as the application will be available. It is a concern when restoring data after years or decades, as it is unlikely that the original application will be available.

Preserving applications (both the underlying database software, and the business application software) over a long period of time will face the same challenges that have already been identified for preserving back-up software. Preserving an image of the software at the time of archiving is problematic as changes to computer systems make it likely that the software eventually will not run. The other option is to install new versions of the software as they are released in order to maintain access. This option assumes that new versions of the software are available; the vendor may cease support of the product, or go out of business entirely. Even if new versions are available, it is likely that, eventually, newer versions will no longer be able to accurately read old data files. In any case, on-going license fees are likely to be payable to vendors.

As an alternative to the use of proprietary database representations, data could be archived in a non-proprietary fashion (e.g. the tables could be expressed as

CSV files). While removing the need for the specific database, it will still be necessary to interpret the stored data to extract information.

3.3. Incremental back-ups

One issue that is not normally a concern when creating a back-up for archival purposes is the use of incremental back-ups. In order to reduce back-up windows and back-up sizes, most back-up strategies perform incremental back-ups. In this, only the data that has changed since the last full back-up is stored. To restore the data it is necessary to first restore the last full back-up, and then each incremental back-up in turn. If this approach was used to create 'archival' back-ups, this would clearly increase the risk of not being able to recover the data. The recovery in this case is clearly a complex process that is more likely to fail than a full back-up. If any one of the incremental back-ups was missing or corrupted, the data could not be recovered.

3.4. Questions:

2. Are you aware of any other issues with using back-up technology to archive data?
3. Do you disagree with the issues we have raised with using back-up technology to archive data?
4. Are you aware of any back-up technology that does not have the limitations identified in this section? Such technology would have to produce backed-up data for which the format was well documented. It would also have to be suitable for archiving data held by Victorian government agencies.
5. Are you aware of any other issues with resurrecting applications in order to access data?
6. Do you disagree with the issues we have raised in resurrecting applications?

4. Issues with using individual pieces of media (e.g. tapes)

Many back-up regimes store the data on off-line media – typically tape, but optical media (DVDs, or CDs), disc drives (removable HDDs), or flash memory could be used. This media is used because 1) it is convenient, 2) it provides a lower cost per megabyte than storing the data on disc drives, and/or 3) the data can easily be moved away from the servers (e.g. stored off-site). Data stored on off-line media takes substantially longer to access than that stored on disc drives, but this in not normally an issue for back-up systems as the data rarely needs to be accessed.

Storage on off-line media for long periods has a number of challenges which will be covered in the following sections. It is possible to address these challenges, and the tasks necessary to do so will be described.

4.1. Media loss

The first problem is the simple loss of the media. Over a long period of time, collections of media are likely to be relocated many times, and preventing pieces of media from being lost during these relocations is a substantial management challenge. Media may also be effectively lost if the knowledge of what is on the media is lost. This can occur if media was not labelled at all, was cryptically or

ambiguously labelled, the media label becomes detached or obliterated, or because the index associating content with media becomes lost.

4.2. Media failure

It is a physical and chemical reality that media will physically deteriorate over time, and the stored data will become harder to read. Examples of physical deterioration include: tape wear, tapes becoming brittle, tape layers sticking together, bearings on HDDs seizing up, heads on HDDs sticking to platters, and chemical changes in recordable optical media. The problem of media failure also includes accidental destruction in readers (e.g. malfunctioning tape readers chewing up tapes).

Physical deterioration is inevitable, and the question is how fast it occurs and the strategies put in place to detect deterioration and replace the media before the data is irretrievable.

Deterioration will occur faster if the media is stored in inappropriate conditions (e.g. higher temperature, higher humidity, dusty or wet environments, near large magnetic fields). Over long periods of storage it is likely that some media will be stored inappropriately, particularly if the media is relocated. Even if the media is stored correctly, the media can deteriorate quickly if the media was originally of poor quality (e.g. poor formulation, or a bad batch).

The only defence against media failure is active management of the media. At least two copies of data should be kept, ideally on two storage technologies, but at least on two brands of media. Media should be periodically refreshed to new media (or, ideally, to new technology). Media should be periodically statistically sampled to determine failure rates, and batches of media that have a higher failure rate should be scheduled for early refresh.

Active management of off-line media for any time is a substantial administrative burden. It should also be noted that reading media (for sampling or refreshing) is a major cause of media failure itself, and procedures will need to be put in place to manage this vulnerability (e.g. when one copy has been destroyed by reading, leaving just one copy which must be read to recover).

4.3. Reader loss

The media readers also physically decay, and after a time will not work reliably. Economics means that new readers will eventually not be available for obsolete technology (this is particularly true for technology that was not common, such as Zip drives). It is expected that modern media technology will be even harder to maintain over a long period due to high levels of miniaturisation. Even if the reader can be repaired, it may not be possible to interface the reader to current computer systems – which may lack the necessary hardware interface, and the software drivers may not work. It is worth noting that physical decay of readers (including mal-adjustment) is likely to be a major cause of media failure. The only effective remedy against reader loss is the periodic refresh of data from one technology to a newer technology. This must, of course, be done before the technology becomes obsolete.

4.4. Back-ups on multiple media

Media loss and media failure is exacerbated if a single back-up is stored on multiple pieces of media. For example, if the back-up is too big to store on one piece of media, or incremental back-ups are used. Loss of any one of the pieces of media will mean loss of the entire back-up.

4.5. Economics

It is recognised that when storing very large quantities of data (e.g. scientific data), the storage costs per gigabyte of tape storage can still cheaper than disc storage.

4.6. Addressing media issues

Issues with media become significant when archiving because of the long periods that media needs to be stored and accessed. Media issues can be addressed, but success requires a long term commitment by an agency to manage the individual media. This requires:

- *Retention of at least two copies*, preferably on different technologies but at least on two brands of media. The copies should be stored in geographically dispersed locations. This needs to be supported by adequate recordkeeping to be able to know, at all times, where all the media is located and what is stored on it.
- *A program of statistically sampling the media* to track the decay of the media, and to detect media batches that are deteriorating faster than expected. This needs to be supported by adequate recordkeeping to be able to identify what media is being held, and what batches individual tapes belong to (or at least its age).
- *An on-going technology watch* to determine when media technology is becoming obsolete. Obsolete means that it is hard (expensive) to source the media and readers or difficult to interface the readers to current computers.
- *A program of refreshing the media* (or normally the technology) when the media starts to deteriorate, or the technology becomes obsolete. As refreshing will require the purchase of new media and hardware, the retrieval and reading of each piece of media, and the updating of records and restoring of the new media, refreshing would be expected to be a major project, and potentially expensive.

4.7. Questions:

7. Are you aware of any other issues with managing individual pieces of media?
8. Do you disagree with the issues we have raised with using individual media to archive data?
9. Can you fore-see any issues with the proposed mechanisms for managing the risk of storing archived data on individual media?

5. Why keep data?

The Victorian public service is required to keep data for the following purposes:

- *To document entitlements*. The government keeps records so that individuals and organisations can demonstrate that they are entitled to certain outcomes. Simple examples are birth records (which show that a person is an Australian citizen) and title records (which show who owns land). More complex examples are police and prison records (which show how members of the public were treated). It is common for critical

reports from the Victorian Auditor General[2] and the Ombudsman[3] to involve records that document entitlements. Records documenting public entitlements are often those that need to be kept for significant periods. Loss of these records has a significant negative impact on individuals or organisations.

- *To support business continuity*. Records document what happened so that future work can be carried out. They also ensure consistency of process and outcomes. This purpose is unlikely to be affected by archiving as data is not normally archived until operational use has ceased. However, some records do not cease to be operational – for example data relating to unsolved crimes
- *To allow agencies to take or defend legal action*. Failure to be able produce data in support of legal cases is likely to have serious financial implications for agencies.

Agencies have a legal obligation to find all relevant documents[4] in the case of legal proceedings (this process is known as 'discovery'). Although this obligation does not extend to heroic efforts to recover data, the cost of discovering documents can be significant.

The Evidence (Miscellaneous Provisions) Act 1958[5] allows judges to make any decision necessary to restore fairness[6] where a document is unavailable[7]. Before making a decision, the court must consider the circumstances in which the document became unavailable and the effect of the unavailability on the case[8].

It is an offence under the Crimes Act 1958[9] to knowingly destroy, or approve the destruction, of data that is, or reasonably likely to be,

[2] For example, the VAGO report into Freedom of Information (http://www.audit.vic.gov.au/reports_and_publications/latest_reports/2011-12/20120418-foi.aspx)

[3] For example, Ombudsman Victoria's own motion investigation into the management and storage of ward records by the Department of Human Services http://www.ombudsman.vic.gov.au/resources/documents/REPORT__Investigation_into_the_storage_and_management_of_ward_records_by_DHS_-_Mar_2012.pdf and into recordkeeping failures by WorkSafe agents http://www.ombudsman.vic.gov.au/resources/documents/New_-_Report_Investigation_into_record_keeping_failures_by_WorkSafe_agents.pdf

[4] Document means any record of information, and includes … (c) anything from which sounds, images, or writings can be reproduced with or without the aid of anything else. (Evidence Act 2008 Definitions)

[5] Part III, Division 9 of the Evidence (Miscellaneous Provisions) Act 1958.

[6] (1) If, in a civil proceeding, it appears to the court that (a) a document is unavailable; and (b) no reproduction of the document is available […]; and (c) the unavailability of the document is likely to cause unfairness to a party to the proceeding – the court […] may make any ruling or order that the court considers necessary to ensure fairness to all parties to the proceeding, having regard to the matter set out in section 89C. (2) Without limiting sub-section (1), a ruling or order may be – (a) that an adverse inference will be drawn from the unavailability of the document; (b) that a fact in issue between the parties be presumed to be true in the absence of evidence to the contrary; (c) that certain evidence not be adduced; (d) that all or part of a defence or statement of claim be struck out; (e) that the evidential burden of proof be reversed in relation to a fact in issue.

[7] A document is unavailable in a civil proceeding if (a) the document is, or has been but is no longer, in the possession, custody or power of a party to the civil proceeding; and (b) the document has been destroyed, disposed of, lost, concealed or rendered illegible, undecipherable, or incapable of identification (s89A).

[8] Before making an order […] the court must have regard to – (a) the circumstances in which the document became unavailable; and (b) the impact of the unavailability on the proceeding, including whether the unavailability of the document will adversely affect the ability of a party to prove its case or make a full defence; and (c) any other matter that the court considers relevant.

[9] Section 254, Crimes Act 1958

required in evidence in a legal proceeding[10]. Corporations can also be convicted of destruction under this Act. Although a conviction under this Act requires a high burden of proof, the penalties that can be imposed are substantial.

- *To support investigations*. To allow internal or external auditors, investigators (e.g. Victoria Police), or regulatory bodies (e.g. Victorian Ombudsman, or the Victorian Auditor-General) to investigate the operations of the agency or crimes. Loss of records is likely to lead to adverse findings and be a political embarrassment.
- *Memory*. To allow Victorians of the future to understand what happened in Victoria and why it occurred. Destruction of records (data) identified as of being permanent value to the State of Victoria is an offence under the *Public Records Act (1973)*[11]. This Act does not distinguish between deliberate destruction, or inadvertent destruction because the record cannot be located or retrieved.

5.1. Questions:

10. Are there any other legal or other risks to agencies if they can no longer access archived data?
11. Do you disagree with the risks we have identified with the loss of archived data?

6. How long does data need to be kept accessible for?

Some data will need to be kept for decades – for example data that uniquely identifies patients must be kept for 75 years after it was last updated[12]. A substantial amount of data in agencies will need to be kept for a decade or more.

Within the Victorian public service there is a legislative regime that governs the decisions about how long data must be retained for (the retention period). These retention periods are set to ensure that citizen's entitlements are protected, that agencies can supply the quality of service they are required to, that agencies are not at undue risk in legal proceedings, that agencies can support investigative activities, and that the Victorian public can understand their history.

Disposal of data is governed by the following principles[13]:

- Disposal of data must be conducted in a lawful manner
- Disposal actions must be based on an informed decision making process

[10] A person who (a) knows that a document or other thing of any kind is, or is reasonably likely to be, required in evidence in a legal proceeding; and (b) either – (i) destroys or conceals it or renders it illegible, undecipherable or incapable of identification; or (ii) expressly, tacitly or impliedly authorises or permits another person to destroy or conceal it or render it illegible, undecipherable or incapable of identification and that other person does so; and (c) acts as described in paragraph (b) with the intention of preventing it from being used in evidence in a legal proceeding – is guilty of an indictable offence and liable to level 6 imprisonment (5 years maximum) or a level 6 fine or both. (Section 254, Crimes Act 1958)

[11] (1) A person who unlawfully removes sells damages or destroys a public record shall be guilty of an offence. (2) Destruction or disposal of public records by a public officer in accordance with standards established under section 12 (i.e. RDAs) is lawful. (Public Record Act (1973) s19)

[12] PROS 11/06 RDA for Patient Information Records Class 1.2.1

[13] Drawn from the Disposal Standard (http://prov.vic.gov.au/government/standards-and-policy/disposal)

- Disposal actions and retention periods for public data must be justifiable

Within the disposal regime, minimum data retention periods[14] are proposed by agencies and are approved by the Keeper of Public Records after a consultation process. The retention periods established are based on legislative requirements, agency administrative requirements, industry practice, and significance of the data for the community beyond its original purpose.

Retention periods can range from 'dispose of once use has ceased', to 'keep indefinitely'. A substantial number of retention periods are less than three years, and many others range from seven to 25 years.

The agreed retention periods are contained in Retention and Disposal Authorities (RDAs)[15] which are available from the PROV website[16].

6.1. Questions:

12. Do you have any comments on this process for determining the period for which data must be kept accessible?

7. Can data be destroyed automatically after seven years?

No, there is no automatic permission to destroy data after seven years. Data retention periods are set by the RDAs – these may be shorter than seven years, but are frequently longer. This misunderstanding may have arisen because a seven-year retention period is common for financial records.

8. Can archived media be disposed of?

If an agency is holding media on which data has been archived, it cannot dispose of that media until the relevant retention period has expired (or the data on the media has been copied to other media).

If an agency does not know exactly what is held on stored media, it must not dispose of that media until the longest possible retention period has expired. This may require agencies to hold the media indefinitely as some retention periods require permanent retention.

It is an agency's responsibility to determine retention periods and authorise disposal. Organisations storing media must not dispose of the media without explicit authorisation from the agency lest they take on legal liability of the disposal.

9. How long is it necessary to keep back-up data?

Where a back-up is solely used for its intended purpose (i.e. to allow for the recovery of data after a failure, error, or disaster), it is only necessary to keep the

[14] Sometimes, data may need to be kept for longer than this minimum period. Data cannot be destroyed, for example, where it is required as evidence in a legal proceeding.

[15] RDAs come in two types. General RDAs authorise the disposal of data that are found in all agencies for example financial and human resources records, or apply to agencies that perform the same function (for example, higher education institutions, local government, schools, and water authorities). Function specific RDAs are issued to authorise the disposal of data and records that is unique to an individual agency.

[16] RDAs are available at http://prov.vic.gov.au/government/disposal-and-transfer/retention-and-disposal-authorities

back-up for as long as it is required[17]. This is because the application itself holds the 'record' of the data.

In particular, it is not necessary to hold back-ups, or the media on which the back-ups are written, for seven years.

10. What are the consequences of destroying data prematurely?

Destruction of data before the expiry of the retention period means that the agency, and its head, is in breach of the *Public Records Act (1973)*. The agency is also at legal risk in legal proceedings. In extreme cases, the person destroying the data, and the managers authorising the destruction, can be jailed or fined. Agencies may be subject to adverse findings by regulatory bodies (e.g. the Victorian Auditor General), or investigatory bodies (e.g. Royal Commissions).

11. References

Management of Backups, Public Records Brief, Queensland State Archives, July 2010, http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/ManagementOfBackups.pdf

Management of Backups, Information Management Advice 25, Tasmanian Archive & Heritage Office, August 2011, http://www.linc.tas.gov.au/__data/assets/pdf_file/0005/341366/TAHO_Advice_-_25.pdf

End of Document

[17] PROS 07/01, General Retention & Disposal Authority for Records of Common Administrative Functions, Class 19.5.3. http://prov.vic.gov.au/wp-content/uploads/2011/05/PROS07-01CommonAdminVar1-WebVersion20110519.pdf