



Public Record Office Victoria  
Advice to Victorian Agencies  
June 2006, Version 2.1

# Advice

# 14

Advice on  
Export of Electronic Records to PROV  
PROS 99/007 (Version 2) Specification 5

Copyright 2003 & 2006, Public Record Office Victoria

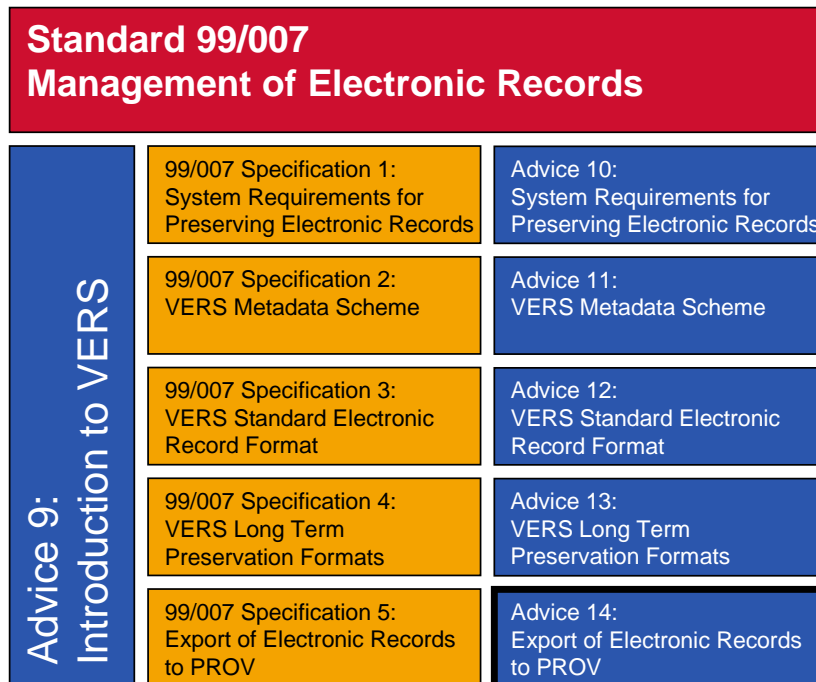
Further copies of this document can be obtained from the PROV Web site  
<http://www.prov.vic.gov.au/>

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 Advice.

Version	Version Date	Details
2.0	31 Jul 03	Released
2.1	30 Jun 06	Updated to reflect digital archive requirements

## The Victorian Electronic Records Strategy (VERS)

This document is a guide to *PROS 99/007 Specification 5: Export of Electronic Records to PROV*. The relationship between the VERS Standard, the Specifications that support this Standard, and the Introduction and Advices that explain VERS is shown below.



These documents have the following purposes:

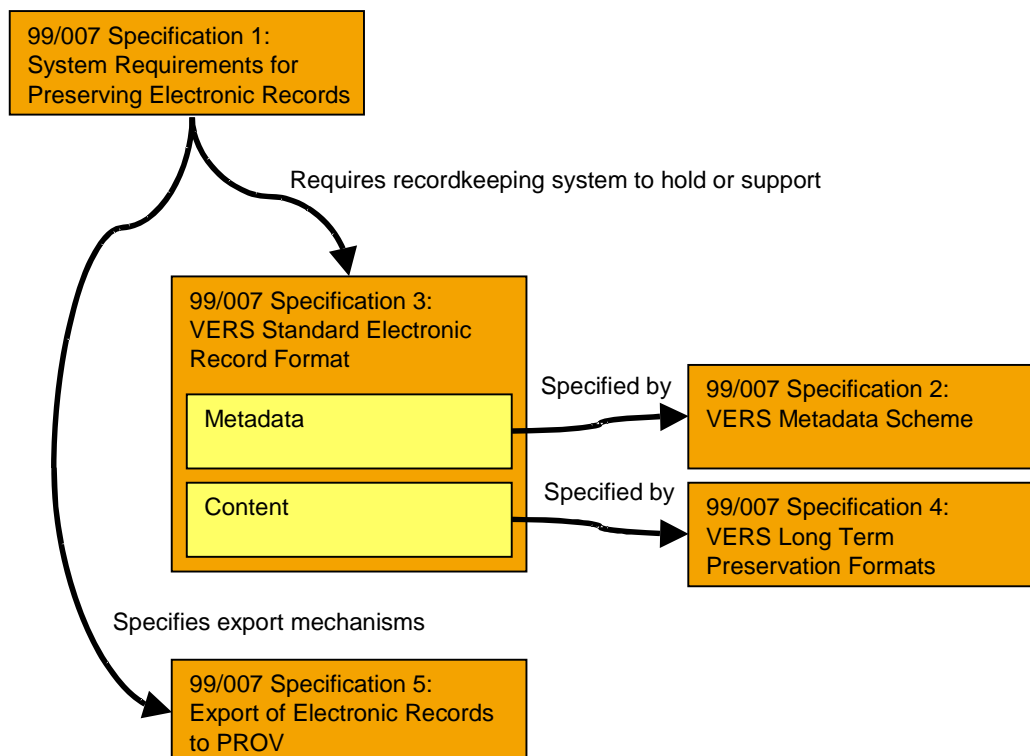
- *Management of Electronic Records*. This document is the Standard itself and is primarily concerned with conformance. The technical requirements of the Standard are contained in five Specifications.
- *Introduction to VERS*. This document provides background information on the goals and the VERS approach to preservation. Nothing in this document imposes any requirements on agencies.
- *Specifications*. These five documents provide the technical requirements that support the Standard. Agencies *must* conform to the mandatory requirements of the specifications, *must* conform to the conditional requirements of the specifications if the appropriate conditions are satisfied, and *may* conform to the optional requirements. Some optional requirements are strongly recommended and these are noted as such.

The five Specifications are:

- *Specification 1: System Requirements for Preserving Electronic Records*. This document specifies the overall functions that a recordkeeping system must perform to preserve electronic records for a substantial period.
- *Specification 2: VERS Metadata Scheme*. This document specifies the metadata that a recordkeeping system must hold to conform to VERS.
- *Specification 3: VERS Standard Electronic Record Format*. This document contains the technical definition of the VERS Encapsulated Object (VEO) format; the mandatory long-term format for records.

- *Specification 4: VERS Long Term Preservation Formats.* This document lists the data formats that PROV accepts as suitable for representing documents for a significant period.
  - *Specification 5: Export of Electronic Records to PROV.* This document lists the approved media and mechanisms by which PROV will accept an export of electronic records.
- *Advices.* These six documents provide background information, explanatory material, and examples in support of the Standard and associated Specifications. None of the information in the Advices imposes any requirement on agencies.

*Relationship between Specifications.* A second view of the relationship between the five Specifications is shown in the following diagram:



*Specification 1 (System Requirements for Preserving Electronic Records)* details the overall requirements on a recordkeeping system for preserving electronic records over a significant period. Amongst other requirements, the recordkeeping system must be capable of exporting the records in a standardised format.

The overall features of this standardised format are defined in *Specification 3 (VERS Standard Electronic Record Format)*, but some details are defined in two other Specifications. *Specification 2 (VERS Metadata Scheme)* defines the meaning and allowed values of the metadata that appears in a record. *Specification 4 (VERS Long Term Preservation Formats)* defines the formats in which the record content must be expressed.

*Specification 5 (Export of Electronic Records to PROV)* defines the mechanisms by which records are exported to PROV.

*Relation to Version 1 of this Standard.* This version of the VERS Standard completely replaces Version 1 of the Standard. Version 2 is identical in its base requirements, but makes those requirements clearer and more explicit. It also contains a number of conditional and optional extensions to Version 1.

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>7</b>
<b>2</b>	<b>Change history .....</b>	<b>7</b>
2.1	Changes between Version 1 and Version 2.0.....	7
2.2	Changes between Version 2.0 and Version 2.1 .....	7
<b>3</b>	<b>Compliance.....</b>	<b>7</b>
<b>4</b>	<b>An overview of exporting to PROV.....</b>	<b>8</b>
4.1	Points to consider .....	8
4.2	Transfers, Series, Consignments, and Sets .....	9
4.3	The export process of a Set.....	10
<b>5</b>	<b>The Manifest .....</b>	<b>13</b>
5.1	An example Manifest for an Internet export .....	16
5.2	An example Manifest for a physical export .....	17
5.3	An example XSL transformation for extracting manifests from VEOs.....	18
<b>6</b>	<b>The Custody Report.....</b>	<b>20</b>
6.1	Custody Acceptance Protocol.....	20
6.2	Example Custody Report .....	23
<b>7</b>	<b>Export over the Internet.....</b>	<b>23</b>
<b>8</b>	<b>Export on Physical Media.....</b>	<b>29</b>
8.1	Transport of Media .....	29
8.2	Labelling of Media .....	30
8.3	Export Media .....	30
8.3.1	CDs .....	31
8.3.2	DVDs.....	32
8.3.3	Tape Media .....	33
8.4	Media Management Software .....	34
8.4.1	Volume management .....	35
8.4.2	Media mounting software .....	35
8.4.3	Media driver software .....	35
8.4.4	Data writing software.....	35
8.4.5	Summary of software requirements .....	36
<b>9</b>	<b>References.....</b>	<b>36</b>



# 1 Introduction

**Information in this Advice is purely informative. Nothing in this Advice imposes any requirements on a VERS compliant implementation; requirements are only imposed by the associated Specifications.**

The purpose of this Advice is to explain the mechanisms defined in the associated Specification to physically export VEOs (primarily electronic records) to PROV. These mechanisms cover the:

- Manifest used to signal to PROV the VEOs included in the set
- Mechanisms used to export the VEOs to PROV (either physical media or the internet).
- Mechanism by which the acceptance of custody of the records represented by the VEOs is indicated back to the agency.

## 2 Change history

### 2.1 Changes between Version 1 and Version 2.0

Specification 5 is new in Version 2.0, however some of the content was formerly in Specification 3. In addition, the following changes have been made:

- The content has been focussed specifically on the export of VEOs to PROV.
- The requirements on media have been made more precise.
- Additional requirements have been specified on the method of representing VEOs on the media.
- The mechanism used to signal that PROV has accepted responsibility for a VEO has been defined.

### 2.2 Changes between Version 2.0 and Version 2.1

Specification 5 has been revised to reflect the changes in the export process due to the implementation of the digital archive at PROV.

## 3 Compliance

Compliance to the associated Specification is achieved by:

- Generation of a Manifest controlling the export of the set.
- Export of VEOs from the source system on one of the media types listed in the Specification or over the Internet. Note that the Specification normally restricts the options for each type of media and these restrictions must be observed.

- If the VEOs are being transferred by means of media, the VEOs must be represented on the media in the file format given in the Specification.
- The agency must accept and process the custody message defined in the Specification (either manually, or automatically using suitable software). Export is not complete until PROV has signalled that it accepts custody of a VEO, and the exporting software must be capable of re-exporting one or more VEOs if it does not receive this indication.

Non compliant exports may be accepted by negotiation with PROV. Non compliant exports may require the purchase of hardware and the development of software by PROV, and so the acceptance will depend on the number of records to be exported, the cost of exporting them, and who will bear the cost.

It should be noted that the examples in this Advice are illustrative of the practice with the digital archive at PROV. The detail may vary with other digital archives using VERS.

## 4 An overview of exporting to PROV

In discussing the export of VEOs to PROV, it should be borne in mind that the export is only a small portion of a larger transfer process. The larger process encompasses, among other tasks:

- negotiation between the agency and PROV as to the records to be transferred and the timing
- determining the series that the record belongs to, and dividing the transfer into series and consignments
- agreeing on the access authority
- performing the export of the records to PROV
- quality control on the received records

This specification deals solely with the export of VEOs from the agency to PROV.

Within the process of exporting VEOs, PROV supports two mechanisms for the movement of VEOs from an agency to PROV. The first is electronic movement over the Internet via the WEBDAV protocol. The second is physical movement by writing the VEOs onto pieces of media (e.g. DVDs or tapes) and physically transferring the media to PROV. The overall export process is similar irrespective of which method is chosen to physically move the VEOs to PROV.

### 4.1 Points to consider

When designing the exporting function for a recordkeeping system, the following points should be considered:

- It is common for PROV to split a transfer into separate Series, and then Series into separate Consignments.

The Series number forms part of the vers:VEOIdentifier element. Consequently, the Series identifier must be assigned before the VEOs can be exported from a recordkeeping system.



VEOs are exported to PROV grouped by Series and Consignment. It is therefore necessary to be able to group the VEOs by Series and Consignment as part of the export process. These groupings are known as 'Sets'.

- VEOs may be generated and exported directly from the recordkeeping system, or they may be generated from the recordkeeping system into a staging area (e.g. disc) and then exported in a second step.
- Manifests may be generated directly by the recordkeeping system, or they may be generated from the VEOs in a staging area.
- Custody reports **MUST** be reconciled with the original records in the recordkeeping system, not with VEOs held in a staging area. Reconciliation against VEOs in a staging area would not detect VEOs that were not properly exported from the recordkeeping system, or were lost from the staging area before the generation of the manifest.

## 4.2 Transfers, Series, Consignments, and Sets

A transfer of electronic records from an agency may cover records on many different topics. Each transfer is broken down into the following groups.

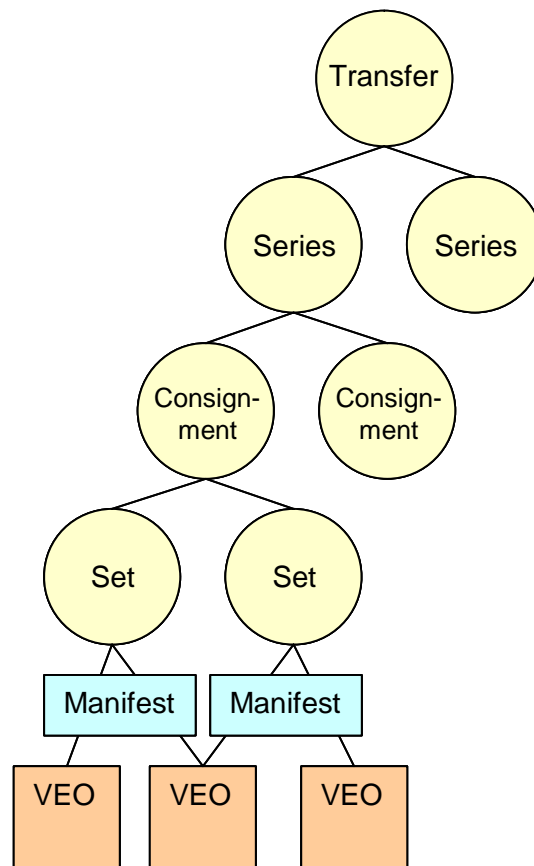


Figure 1. The breakdown of a Transfer into Series, Consignments, Sets, and VEOs.

A transfer consists of one or more *Series*. A series is a collection of related records. The division of a transfer into series will be carried out by an archivist as part of the broader transfer process. The Series identification forms part of the unique identifier of each VEO. Separation of the records into Series must consequently be carried out before the VEOs can be exported from a recordkeeping system.

A Series consists of one or more Consignments. A consignment is a subgrouping of records within a Series. Again the division of a series into consignments will be carried out by an archivist as part of the broader transfer process. The Consignment identifier is not contained within the VEOs, but is included in the manifest.

A Consignment consists of one or more Sets. Sets are the key management tool within the digital archive. All the VEOs in one Set are processed together (although individual VEOs may be deleted from the Set during processing). Sets may be used in a variety of ways:

- To break a very large Consignment into Sets of a convenient size for the source record keeping system or digital archive.
- To separate different VEOs, if this is required. For example, File VEOs may be separated from Record VEOs. Alternatively, very large VEOs may be separated out from smaller VEOs.
- To resubmit VEOs that were not accepted by the digital archive. This is discussed further below.

A Set contains one or more VEOs. Normally, a VEO is only contained in one Set. However, in some circumstances the same VEO will be present in several Sets (e.g. when resubmitting VEOs). If the same VEO is present in multiple Sets, the digital archive will detect and discard the duplicates.

The VEOs contained in a Set are listed in a Manifest (see section 5). The Manifest is an XML document generated by the agency performing the transfer. It is used to prepare the digital archive to accept the set. The Manifest identifies the Series and Consignment that the Set belongs to. Consequently, the Manifest must be generated after the Series and Consignments have been identified. Where the set is to be transferred using physical media (e.g. CDs), the Manifest includes a list of media. In this case the Manifest cannot be generated until the VEOs have been divided amongst the media.

### 4.3 The export process of a Set

The overall export process of a Set is tabulated below. This export process is the same irrespective of whether the movement of VEOs is via the Internet or by physical media.

#### Agency Action

- The records manager at the transferring agency creates the VEOs and Manifest.
- The Manifest is sent via email to the transfer archivist at PROV

#### PROV Action

- The transfer archivist loads the Manifest into the digital archive where it is used to prepare the digital archive to receive the Set. Loading the Manifest, for example, sets up a transfer workflow that will contain the Set during processing. It also creates an inbox for the Set into which the VEOs will be placed. A Set name and password for the inbox is generated.
- The transfer archivist sends an acknowledgement to the records manager, normally by email. The acknowledgement includes the Set name and, if the transfer is

to be over the Internet, the URL of the digital archive, the details of the inbox and the password.

- The records manager transfers the VEOs to PROV.
  - In the case of a physical transfer, the VEOs are written to one of the approved types of media using the TAR or PAX format, and the media is sent to PROV.
  - In the case of an Internet transfer, the VEOs are transferred over the Internet to the inbox using the WebDAV protocol. The records manager logs onto the digital archive using the URL of the digital archive, the Set name, and the password. When all of the VEOs have been loaded into the inbox, the records manager loads a special file named the 'end-of-set.trigger' file into the inbox to start the processing.
    - In the case of a physical transfer, the transfer archivist receives the media and loads the VEOs into the inbox. The archivist also loads the 'end-of-set.trigger' file.
    - The digital archive commences processing of the Set once the end-of-set trigger file is detected in the inbox. Processing commences with the VEOs being moved from the inbox to the digital archive itself
    - Once the VEOs are in the digital archive, the following checks are carried out:
      - Reconciliation against the Manifest. A report is produced listing the VEOs listed in the Manifest but which have not been received, and those VEOs that have been received, but not are not listed in the Manifest.
      - Validity against the VERS standard. VEOs that fail this check are deleted from the Set. A report is produced of VEOs that fail the validity check. It is the responsibility of the agency to fix the problem and resubmit the failed VEOs as a new Set.
      - Virus infections. Infected VEOs are deleted from the Set. A report is produced of VEOs that fail the virus check. It is the responsibility of the agency to fix the problem and resubmit the failed VEOs as a new Set.
    - If too many VEOs in the Set have failed the checks, the transfer archivist may decide to abort the transfer. In this case the records manger will be notified.
    - The VEOs remain in quarantine for a period of time, currently one week. At the end of this time, the VEOs are checked for viruses a

- second time. Again, infected VEOs are deleted from the set and it is the agencies responsibility to fix the problem and resubmit the failed VEOs as a new Set. If too many VEOs in the Set are infected, the transfer archivist may decide to abort the transfer.
  - After the VEOs come out of quarantine, the transfer archivist will check the VEOs to ensure that they are the records that were agreed to be transferred. Incorrect records will be deleted. The transfer archivist again has the option to abort the transfer.
  - The transfer archivist will authorise the VEOs to be moved from the processing area to the permanent store.
  - PROV takes custody of the digital records once the VEOs have been confirmed as being accurately copied to the permanent store, with a second copy accurately copied to off-site storage. The digital archive will produce a custody report listing all VEOs that have been accepted into the digital repository. The custody report is sent to the records manager
- When the custody report is received, the transfer of this Set is complete. This Custody Report is the official notification that PROV has accepted custody of specific VEOs. The Custody Report lists each VEOs in the set for which PROV has accepted custody. The records manager can destroy the agency copy of any record listed in the custody report. Note, however, that some records originally copied to the digital archive may not appear in the custody report. These records may have been infected with a virus, not be valid VEOs, or not listed in the manifest.

If a VEO is not listed in the Custody Report, PROV has not accepted custody of it (typically because it was invalid, virus infected, or not permanent). It is the agencies responsibility to:

- identify those VEOs in the Set that have not been accepted
- correct (if necessary) the problem that caused failure
- resubmit the fixed VEOs as a new Set within the Transfer

It is expected that a compliant recordkeeping system will support the agency in carrying out these responsibilities.

The cycle of receiving a Custody Report, fixing the problems, and creating a new Set continues until all the records that form part of the Transfer have been successfully transferred to PROV.

Once PROV has accepted custody of a record, the agency is free to delete its copy of the record.

## 5 The Manifest

Each set exported to PROV must be preceded by a manifest. Each Set exported to PROV has an associated Manifest which lists every VEO in the Set. The Manifest contains a small amount of information about each VEO.

PROV uses the manifest in two ways:

- It is loaded into the digital archive to manage the processing of the Set. The information in the manifest is used to track, and report on, the progress of every VEO in the set.
- It is used by the transfer archivist to perform preliminary quality assurance on the transferred set. This quality assurance checks whether the VEOs are what was agreed to be accepted. This cannot be done directly from the VEOs as the archivist cannot access the contents of a VEO until they have completed quarantine in the digital archive.

A manifest consists of three sections. The first two are mandatory, while the third only appears in manifests describing a physical transfer.

The first section identifies the set. The information contained in this section consists of:

- `dam:created_timestamp`. This is the date and time the manifest was created. The date/time is expressed in the form defined in PROS 99/007 (Version 2), Specification 2, section 14. An example is “2005-07-13T11:35:07-10:00”
- `dam:agency_identifier`. This is the VA number, without the leading “VA”. An example is “473”
- `dam:series_type` (always the string ‘VPRS’)
- `dam:series_number`. This is the VPRS number, without the leading “VPRS”. An example is “12800”
- `dam:job_id`. This is the PROV job identifier. It is allocated by PROV, an example is “TR 2005/0021”
- `dam:consignment_type`. This is usually the string “P”
- `dam:consignment_number`. This must be four digits long and be padded with leading zeros if necessary. An example is “0001”

An example first section is:

```
<dam:created_timestamp>2005-07-13T11:35:07-10:00</dam:created_timestamp>
<dam:agency_id>473</dam:agency_id>
<dam:series_type>VPRS</dam:series_type>
<dam:series_number>110</dam:series_number>
<dam:job_id>TR 2005/0021</dam:job_id>
<dam:consignment_type>P</dam:consignment_type>
<dam:consignment_number>0001</dam:consignment_number>
```

The second section lists the VEOs. It consists of a sequence of `dam:manifest_object_item`. Each of these describes one VEO. The information contained in this section consists of:

- `dam:computer_filename`. The filename of the VEO on the media or when transferred over the internet, for example “00110-P0000-000001.veo”. This element is used by the digital archive to load the VEO into the transfer process. Note that this is not the pathname of the VEO (i.e. it does not include the directories from the root of the file system). The filename is limited to 256 characters long.

- `dam:file_identifier`. This is the contents of the `vers:FileIdentifier` (M102) element in the VEO. This element is only used by the transfer archivist to validate the transfer. An example would be “Applications 1995”. The value has a maximum length of 15 characters, and if the `vers:FileIdentifier` element is longer than this it should be truncated at the beginning or end to this length.
- `dam:vers_record_identifier`. This is the contents of the `vers:VERSRecordIdentifier` (M103) element in the VEO. This element is only used by the transfer archivist to validate the transfer. An example would be “00110-P0000-000001”. If the VEO does not have a `vers:VERSRecordIdentifier` (i.e. is a FileVEO), the element must be present, but empty (i.e. `<dam:vers_record_identifier xsi:nil="true"/>`). The value has a maximum length of 15 characters, and if the `vers:VERSRecordIdentifier` element is longer than this it should be truncated at the beginning or end to this length.
- `dam:veo_title`. This is the contents of the `vers:TitleWords` (M35) element. This element is only used by the transfer archivist to validate the transfer. An example would be “Volume 1: Register of Convicts: 1842 – 1854”. The value has a maximum length of 1024 characters, and if the `vers:TitleWords` element is longer than this it should be truncated at the beginning or end to this length.
- `dam:veo_classification`. This contains the classification of the VEO. The classification is obtained from the `naa:Function` (M50) element or the `vers:Subject` (M37) element. If neither is present, a default value is generated. This element is only used by the transfer archivist to validate the transfer. The value is generated as follows:
  - if the `naa:Function` (M50) element is present, the value is the concatenation of `naa:FunctionDescriptor` (M51), `naa:ActivityDescriptor` (M52), and `naa:ThirdLevelDescriptor` (M53) with spaces between. For example: `<dam:veo_classification>Record Keeping Standards VERS 2006 Revision</dam:veo_classification>`
  - If the `vers:Subject` (M37) element is present, the value is the concatenation of “(”, the `vers:KeywordLevel` (M38), “ ”, the `vers:Keyword` (M39), the `vers:Subject` (recursive), and “)”. For example: `<dam:veo_classification>(1 FirstLevelSubject (2 2ndLevelSubject)) </dam:veo_classification>`
  - If neither `naa:Function` (M50) or `vers:Subject` (M37) element is present, the value is set to “No classification”. For example: `<dam:veo_classification>No classification</dam:veo_classification>`

It has a maximum length of 1024 characters; if the value is longer than this it must be truncated at the beginning or end to this length.

- `dam:veo_access_category`. This contains the value of the `naa:AccessStatus` (M29) element, if present, otherwise the text “Not specified”. Examples are: `<dam:veo_access_category>Not for Release</dam:veo_access_category>` or `<dam:veo_access_category>Not specified</dam:veo_access_category>`. This element is only used by the transfer archivist to validate the transfer. It has a maximum length of 1024 characters, and must be truncated to this length if longer.
- `dam:veo_disposal_authority`. This contains the value of the `naa:Sentence` (M90) element. This element is only used by the transfer archivist to validate the transfer. It has a maximum length of 1024 characters, and must be truncated to this length if longer.
- `dam:veo_date_range`. This contains the date range that the VEO covers. The date range is delimited by the start and end dates. This element is only used by the transfer archivist to validate the transfer.
- `dam:veo_start_date`. This contains the value of the `naa:DateTimeRegistered` (M57) element. The date/time is expressed in the form defined in PROS 99/007 (Version 2), Specification 2, section 14. An example is “2005-07-13T11:35:07-10:00”

- `dam:veo_end_date`. In the case of a record VEO, or a File VEO that has not been closed, this element is empty (i.e. `<dam:veo_end_date xsi:nil="true"/>`). In the case of a File VEO that has been closed, the element contains the value of the `vers:DateTimeClosed` (M144) element. The date/time is expressed in the form defined in PROS 99/007 (Version 2), Specification 2, section 14. An example is “2005-07-13T11:35:07-10:00”.
- `dam:size_kb`. This contains the size of the VEO in kilobytes (1024 bytes). This element is only used by the transfer archivist to validate the transfer. The maximum size is 999000000 kB.

An example second section for a record VEO is:

```
<dam:manifest_object_item>
  <dam:computer_filename>00110-P0000-000001.veo</dam:computer_filename>
  <dam:file_identifier>8164556315</dam:file_identifier>
  <dam:vers_record_identifier>
    00110-P0000-000001
  </dam:vers_record_identifier>
  <dam:veo_title>
    Volume 1: Register of Convicts: 1842 - 1854
  </dam:veo_title>
  <dam:veo_classification>No classification</dam:veo_classification>
  <dam:veo_access_category>Not specified</dam:veo_access_category>
  <dam:veo_disposal_authority>
    No Disposal Coverage
  </dam:veo_disposal_authority>
  <dam:veo_date_range>
    <dam:veo_start_date>2005-06-20T16:38:15-10:00</dam:veo_start_date>
    <dam:veo_end_date xsi:nil="true"/>
  </dam:veo_date_range>
  <dam:size_kb>221370</dam:size_kb>
</dam:manifest_object_item>
```

The third section lists the media (e.g. CDs) on which the VEOs are being transferred to PROV. This section is only present for a physical transfer. The information contained in this section consists of a sequence of `dam:media_item` elements, one for each piece of media in this set. The elements within the `dam:media_item` element are:

- `dam:media_written_date`. This is the date the media was written. The date/time is expressed in the form defined in PROS 99/007 (Version 2), Specification 2, section 14. An example is “2005-07-13”. This element is used to manage the media within the digital archive.
- `dam:media_item_number`. This is the sequential number of the piece of media within the collection of media for this set. An example value is “1”. This element is used to manage the media within the digital archive.
- `dam:media_item_total_number`. This is the total number of pieces of media used to export this set. An example value is “4”. This element is used to manage the media within the digital archive.
- `dam:media_type`. This is the type of media. Valid values are “CD”, “DVD”, “LTO”, and “DDS”.

An example second section for a record VEO is:

```
<dam:media_list>
  <dam:media_item>
    <dam:media_written_date>2004-09-30</dam:media_written_date>
    <dam:media_item_number>1</dam:media_item_number>
    <dam:media_item_total_number>4</dam:media_item_total_number>
```

```

    <dam:media_type>DVD</dam:media_type>
  </dam:media_item>
</dam:media_list>

```

## 5.1 An example Manifest for an Internet export

The following manifest describes an export where the VEOs are transferred across the internet. It contains two VEOs; a file and a record VEO.

```

<?xml version="1.0" encoding="UTF-8" ?>
<dam:set_manifest xmlns:dam="http://www.prov.vic.gov.au/digitalarchive/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.prov.vic.gov.au/digitalarchive/
  http://www.prov.vic.gov.au/digitalarchive/setManifest_1_0_0.xsd">
  <dam:electronic_transfer>
    <dam:created_timestamp>2005-07-13T11:35:07-10:00</dam:created_timestamp>
    <dam:agency_id>473</dam:agency_id>
    <dam:series_type>VPRS</dam:series_type>
    <dam:series_number>110</dam:series_number>
    <dam:job_id>TR 2005/0021</dam:job_id>
    <dam:consignment_type>P</dam:consignment_type>
    <dam:consignment_number>0001</dam:consignment_number>
    <dam:manifest_object_list>
      <dam:manifest_object_item>
        <dam:computer_filename>
          00110-P0000-000001.veo
        </dam:computer_filename>
        <dam:file_identifier>8164556315</dam:file_identifier>
        <dam:vers_record_identifier>
          00110-P0000-000001</dam:vers_record_identifier>
        <dam:veo_title>
          Volume 1: Register of Convicts: 1842 - 1854
        </dam:veo_title>
        <dam:veo_classification>No classification</dam:veo_classification>
        <dam:veo_access_category>Not specified</dam:veo_access_category>
        <dam:veo_disposal_authority>
          No Disposal Coverage
        </dam:veo_disposal_authority>
        <dam:veo_date_range>
          <dam:veo_start_date>2005-06-20T16:38:15-10:00</dam:veo_start_date>
          <dam:veo_end_date xsi:nil="true"/>
        </dam:veo_date_range>
        <dam:size_kb>221370</dam:size_kb>
      </dam:manifest_object_item>
      <dam:manifest_object_item>
        <dam:computer_filename>
          F00110-P0000-000001.veo
        </dam:computer_filename>
        <dam:file_identifier>8164556315</dam:file_identifier>
        <dam:vers_record_identifier xsi:nil="true"/>
        <dam:veo_title>
          Volume 1: Register of Convicts: 1842 - 1854
        </dam:veo_title>
        <dam:veo_classification>No classification</dam:veo_classification>
        <dam:veo_access_category>Not specified</dam:veo_access_category>
        <dam:veo_disposal_authority>
          No Disposal Coverage
        </dam:veo_disposal_authority>
        <dam:veo_date_range>
          <dam:veo_start_date>2005-06-20T16:37:38-10:00</dam:veo_start_date>
          <dam:veo_end_date>2005-06-20T16:37:38-10:00</dam:veo_end_date>

```



```

    </dam:veo_date_range>
    <dam:size_kb>12</dam:size_kb>
  </dam:manifest_object_item>
</dam:manifest_object_list>
</dam:electronic_transfer>
</dam:set_manifest>

```

## 5.2 An example Manifest for a physical export

The following manifest describes an export where the VEOs are transferred using a CD. It contains two VEOs; a file and a record VEO. It should be noted that the description of the VEOs is identical to that given in the previous example. The only difference is

- The use of the `dam:media_transfer` element instead of the `dam:electronic_transfer` element
- The presence of a `dam:media_list` element

```

<?xml version="1.0" encoding="UTF-8" ?>
<dam:set_manifest xmlns:dam="http://www.prov.vic.gov.au/digitalarchive/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.prov.vic.gov.au/digitalarchive/
  http://www.prov.vic.gov.au/digitalarchive/setManifest_1_0_0.xsd">
  <dam:media_transfer>
    <dam:created_timestamp>2005-07-13T11:35:07-10:00</dam:created_timestamp>
    <dam:agency_id>473</dam:agency_id>
    <dam:series_type>VPRS</dam:series_type>
    <dam:series_number>110</dam:series_number>
    <dam:job_id>TR 2005/0021</dam:job_id>
    <dam:consignment_type>P</dam:consignment_type>
    <dam:consignment_number>0001</dam:consignment_number>
    <dam:manifest_object_list>
      <dam:manifest_object_item>
        <dam:computer_filename>
          00110-P0000-000001.veo
        </dam:computer_filename>
        <dam:file_identifier>8164556315</dam:file_identifier>
        <dam:vers_record_identifier>
          00110-P0000-000001</dam:vers_record_identifier>
        <dam:veo_title>
          Volume 1: Register of Convicts: 1842 - 1854
        </dam:veo_title>
        <dam:veo_classification>No classification</dam:veo_classification>
        <dam:veo_access_category>Not specified</dam:veo_access_category>
        <dam:veo_disposal_authority>
          No Disposal Coverage
        </dam:veo_disposal_authority>
        <dam:veo_date_range>
          <dam:veo_start_date>2005-06-20T16:38:15-10:00</dam:veo_start_date>
          <dam:veo_end_date xsi:nil="true"/>
        </dam:veo_date_range>
        <dam:size_kb>221370</dam:size_kb>
      </dam:manifest_object_item>
      <dam:manifest_object_item>
        <dam:computer_filename>
          F00110-P0000-000001.veo
        </dam:computer_filename>
        <dam:file_identifier>8164556315</dam:file_identifier>
        <dam:vers_record_identifier xsi:nil="true"/>
        <dam:veo_title>
          Volume 1: Register of Convicts: 1842 - 1854
        </dam:veo_title>

```

```

<dam:veo_classification>No classification</dam:veo_classification>
<dam:veo_access_category>Not specified</dam:veo_access_category>
<dam:veo_disposal_authority>
  No Disposal Coverage
</dam:veo_disposal_authority>
<dam:veo_date_range>
<dam:veo_start_date>2005-06-20T16:37:38-10:00</dam:veo_start_date>
<dam:veo_end_date>2005-06-20T16:37:38-10:00</dam:veo_end_date>
</dam:veo_date_range>
<dam:size_kb>12</dam:size_kb>
</dam:manifest_object_item>
</dam:manifest_object_list>
<dam:media_list>
<dam:media_item>
  <dam:media_written_date>2004-09-30</dam:media_written_date>
  <dam:media_item_number>1</dam:media_item_number>
  <dam:media_item_total_number>1</dam:media_item_total_number>
  <dam:media_type>DVD</dam:media_type>
</dam:media_item>
</dam:media_list>
</dam:media_transfer>
</dam:set_manifest>

```

### 5.3 An example XSL transformation for extracting manifests from VEOs

The following XSL transformation extracts a `dam:manifest_object_item` from a VEO. Note that it does not limit the length of the elements.

```

<?xml version="1.0"?>
<xsl:stylesheet
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:dam="http://www.prov.vic.gov.au/digitalarchive/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:naa="http://www.naa.gov.au/recordkeeping/control/rkms/contents.html"
  xmlns:vers="http://www.prov.vic.gov.au/gservice/standard/pros99007.htm"
  exclude-result-prefixes="xsl xsi dam naa vers"
  version="1.0">
  <xsl:output method="xml" omit-xml-declaration="yes" indent="no"
    cdata-section-elements="dam:veo_classification dam:file_identifier
dam:vers_record_identifier dam:veo_title"/>

  <xsl:template match="/">
    <xsl:apply-templates
select="vers:VERSEncapsulatedObject/vers:SignedObject/vers:ObjectContent"/>
  </xsl:template>

  <xsl:template match="vers:ObjectContent">
    <xsl:apply-templates select="vers:Record/vers:RecordMetadata"/>
    <xsl:apply-templates select="vers:File/vers:FileMetadata"/>
    <xsl:apply-templates
select="vers:ModifiedVEO/vers:RevisedVEO/vers:SignedObject/vers:ObjectContent"/>
  </xsl:template>

  <xsl:template match="vers:RecordMetadata|vers:File/vers:FileMetadata">
    <xsl:apply-templates select="vers:VEOIdentifier"/>
    <xsl:choose>
      <xsl:when test="naa:Title">
        <xsl:apply-templates select="naa:Title/naa:TitleWords"/>
      </xsl:when>
      <xsl:otherwise>
        <dam:veo_title>
          No Title
        </dam:veo_title>
      </xsl:otherwise>
    </xsl:choose>
  </xsl:template>

```

```

    </xsl:choose>
  <xsl:choose>
    <xsl:when test="naa:Function">
      <dam:veo_classification>
        <xsl:apply-templates select="naa:Function"/>
      </dam:veo_classification>
    </xsl:when>
    <xsl:when test="vers:Subject">
      <dam:veo_classification>
        <xsl:apply-templates select="vers:Subject"/>
      </dam:veo_classification>
    </xsl:when>
    <xsl:otherwise>
      <dam:veo_classification>
        No classification
      </dam:veo_classification>
    </xsl:otherwise>
  </xsl:choose>
  <dam:veo_access_category>
    Not specified
  </dam:veo_access_category>
  <xsl:apply-templates select="naa:Disposal/naa:Sentence"/>
  <xsl:apply-templates select="naa:Date"/>
  <xsl:apply-templates select="vers:Date"/>
</xsl:template>

<xsl:template match="vers:ModifiedVEO">
  <xsl:apply-templates
select="vers:RevisedVEO/vers:SignedObject/vers:ObjectContent"/>
</xsl:template>

<xsl:template match="vers:VEOIdentifier">
  <xsl:apply-templates select="vers:FileIdentifier"/>
  <xsl:choose>
    <xsl:when test="vers:VERSRecordIdentifier">
      <xsl:apply-templates select="vers:VERSRecordIdentifier"/>
    </xsl:when>
    <xsl:otherwise>
      <dam:vers_record_identifier xsi:nil="true"/>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>

<xsl:template match="vers:FileIdentifier">
  <dam:file_identifier><xsl:apply-templates
select="vers:Text"/></dam:file_identifier>
</xsl:template>

<xsl:template match="vers:VERSRecordIdentifier">
  <dam:vers_record_identifier><xsl:apply-templates
select="vers:Text"/></dam:vers_record_identifier>
</xsl:template>

<xsl:template match="naa:TitleWords">
  <dam:veo_title><xsl:value-of select="normalize-space(.)"/></dam:veo_title>
</xsl:template>

<xsl:template match="naa:Function">
  <xsl:value-of select="naa:FunctionDescriptor"/>
  <xsl:value-of select="naa:ActivityDescriptor"/>
  <xsl:value-of select="naa:ThirdLevelDescriptor"/>
</xsl:template>

<xsl:template match="vers:Subject">
  (
    <xsl:value-of select="normalize-space(vers:KeywordLevel)"/>
    <xsl:text> </xsl:text>
    <xsl:apply-templates select="vers:Keyword"/>
    <xsl:apply-templates select="vers:Subject"/>
  )
</xsl:template>

<xsl:template match="vers:Keyword">

```

```

        <xsl:value-of select="normalize-space(.)"/>
</xsl:template>

<xsl:template match="vers:RightsManagement">
  <xsl:choose>
    <xsl:when test="naa:AccessStatus">
      <dam:veo_access_category>
        <xsl:apply-templates select="naa:AccessStatus"/>
      </dam:veo_access_category>
    </xsl:when>
    <xsl:otherwise>
      <dam:veo_access_category>
        Not specified
      </dam:veo_access_category>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>

<xsl:template match="naa:Sentence">
  <dam:veo_disposal_authority><xsl:value-of select="normalize-
space(.)"/></dam:veo_disposal_authority>
</xsl:template>

<xsl:template match="naa:Date">
  <dam:veo_date_range>
    <dam:veo_start_date><xsl:value-of select="normalize-
space(naa:DateTimeRegistered)"/></dam:veo_start_date>
    <dam:veo_end_date xsi:nil="true"/>
  </dam:veo_date_range>
</xsl:template>

<xsl:template match="vers:Date">
  <dam:veo_date_range>
    <dam:veo_start_date><xsl:value-of select="normalize-
space(naa:DateTimeRegistered)"/></dam:veo_start_date>
    <xsl:choose>
      <xsl:when test="vers:DateTimeClosed">
        <dam:veo_end_date><xsl:value-of select="normalize-
space(vers:DateTimeClosed)"/></dam:veo_end_date>
      </xsl:when>
      <xsl:otherwise>
        <dam:veo_end_date xsi:nil="true"/>
      </xsl:otherwise>
    </xsl:choose>
  </dam:veo_date_range>
</xsl:template>

<xsl:template match="vers:Text">
  <xsl:value-of select="normalize-space(.)"/>
</xsl:template>
</xsl:stylesheet>

```

## 6 The Custody Report

Please note that in earlier versions of this Specification the 'Custody Report' was known as the 'Acceptance Message'. It was renamed in version 2.1 to emphasise that it signals acceptance of custody of the records by PROV. The XML element, however, has not been renamed.

### 6.1 Custody Acceptance Protocol

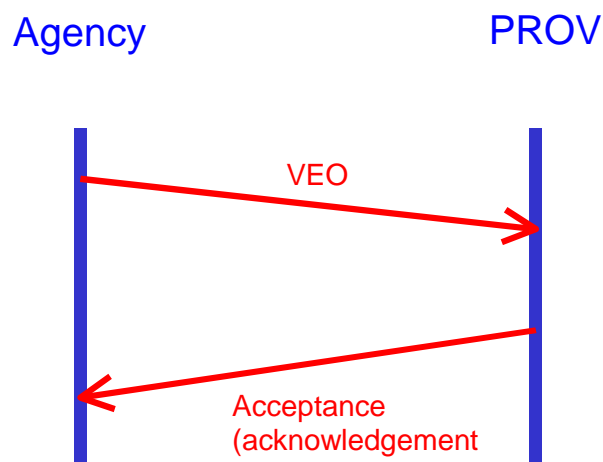
A key part of the process of exporting VEOs is the transfer of responsibility for the preservation of the records. It is essential that the agency exporting the records can clearly

identify when the responsibility for preserving the records has been accepted by PROV so that the agency can discard the original copies.

The process of exporting VEOs to PROV can be considered as a transmission protocol transmitting VEOs from the exporting system to PROV system. Accordingly we have adapted a positive acknowledgement transmission protocol to the export of records.

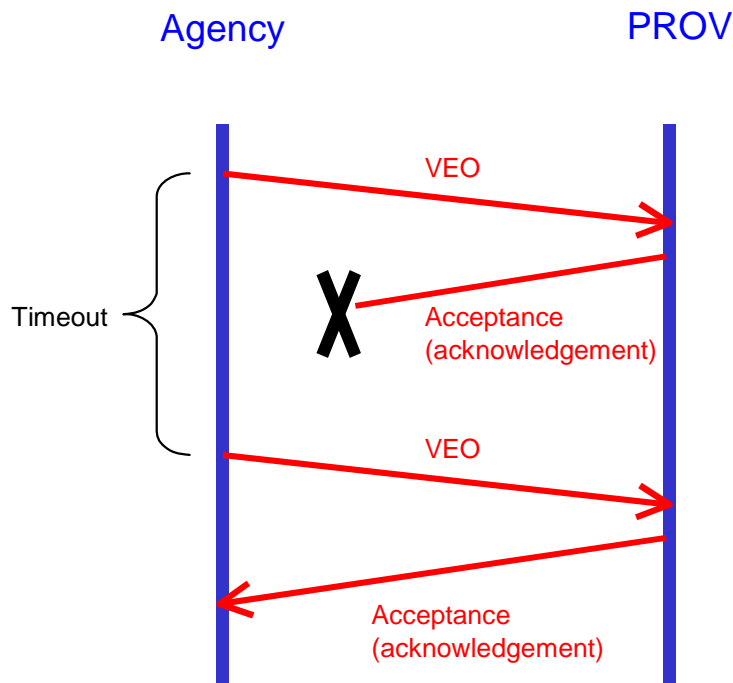
In this approach PROV responds with an acknowledgement when it has accepted responsibility for a record. We refer to this acknowledgement as a 'custody report' to emphasise the meaning of the message.

The normal operation of this process is shown in the following diagram. The agency sends the VEO to PROV. After PROV has accepted responsibility for the contents of the VEO it sends a custody report back to the agency. The agency can then discard its copy of the VEO. It is important to note that the custody report is an acceptance of responsibility. It is not just an acknowledgement of receipt.



PROV may accept responsibility for a number of VEOs in one custody report.

If something goes wrong with the export, the operation of the protocol is shown in the following diagram. In this example, the custody report has been lost in the transfer back to the agency. Because the agency has not received the custody report within a specific period of time (known as a 'timeout'), it resends the VEO. PROV reprocesses the VEO and sends a new custody report.



The timeout period is not defined in the Specification as it will depend on the processing time within PROV and on the mechanism used to transfer information between the agency and PROV. The re-transmission of the VEO may be triggered automatically or manually.

This protocol has both benefits and costs.

The major benefit is that the process is self-healing. In the event of an error in transferring the VEO, the system will recover. Failures that will be corrected include:

- The agency believing that they had sent the VEO when they had not.
- Loss of the VEO during the export to PROV.
- Loss of the VEO at PROV.
- PROV not transmitting the custody report.
- Loss of the custody report during its transmission to the agency.
- Loss of the custody report by the agency.

However, the implications of this protocol are that:

- The agency must keep track of the VEOs that it has sent to PROV and resend them if no custody report is received within the specified period.
- The agency must mark VEOs that have been accepted so that they are not resent.
- PROV must be capable of accepting multiple copies of the same VEO without error and resending a custody report if it has received them before.
- The agency must implement the protocol.

## 6.2 Example Custody Report

The following is an example of a minimal acceptance message that accepts custody of two VEOs. Note that in earlier versions of this specification, the custody report was known as the 'acceptance message'.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE vers:AcceptanceMessage
  SYSTEM "http://www.prov.vic.gov.au/vers/standard/vers.dtd">
<vers:AcceptanceMessage
  xmlns:vers="http://www.prov.vic.gov.au/gservice/standard/pros99007.htm">
  <vers:Version>1.0</vers:Version>
  <vers:TransferReference>TR 0003/2006</vers:TransferReference>
  <vers:AcceptanceDate>2006-02-17T19:35:16-10:00</vers:AcceptanceDate>
  <vers:Acknowledgement>
    <vers:YourReference>
      <vers:VEOIdentifier>
        <vers:AgencyIdentifier>
          <vers:Text>2876</vers:Text>
        </vers:AgencyIdentifier>
        <vers:SeriesIdentifier>
          <vers:Text>12800</vers:Text>
        </vers:SeriesIdentifier>
        <vers:FileIdentifier>
          <vers:Text>1164095484</vers:Text>
        </vers:FileIdentifier>
      </vers:VEOIdentifier>
    </vers:YourReference>
  </vers:Acknowledgement>
  <vers:Acknowledgement>
    <vers:YourReference>
      <vers:VEOIdentifier>
        <vers:AgencyIdentifier>
          <vers:Text>2876</vers:Text>
        </vers:AgencyIdentifier>
        <vers:SeriesIdentifier>
          <vers:Text>12800</vers:Text>
        </vers:SeriesIdentifier>
        <vers:FileIdentifier>
          <vers:Text>1164095484</vers:Text>
        </vers:FileIdentifier>
        <vers:VERSRecordIdentifier>
          <vers:Text>12800-00001-000030-010</vers:Text>
        </vers:VERSRecordIdentifier>
      </vers:VEOIdentifier>
    </vers:YourReference>
  </vers:Acknowledgement>
</vers:AcceptanceMessage>
```

## 7 Export over the Internet

PROV will accept records exported using the WebDAV protocol [WEBDAV]. The digital archive is WebDAV Class 1 compliant.

Important points to note in transferring VEOs to PROV over the internet are:

- The URL of the inbox is [http://prov.vic.gov.au/sets/<set\\_id>](http://prov.vic.gov.au/sets/<set_id>), where the <set\_id> is the identifier of the set. The set identifier will be given to you by the transfer archivist.

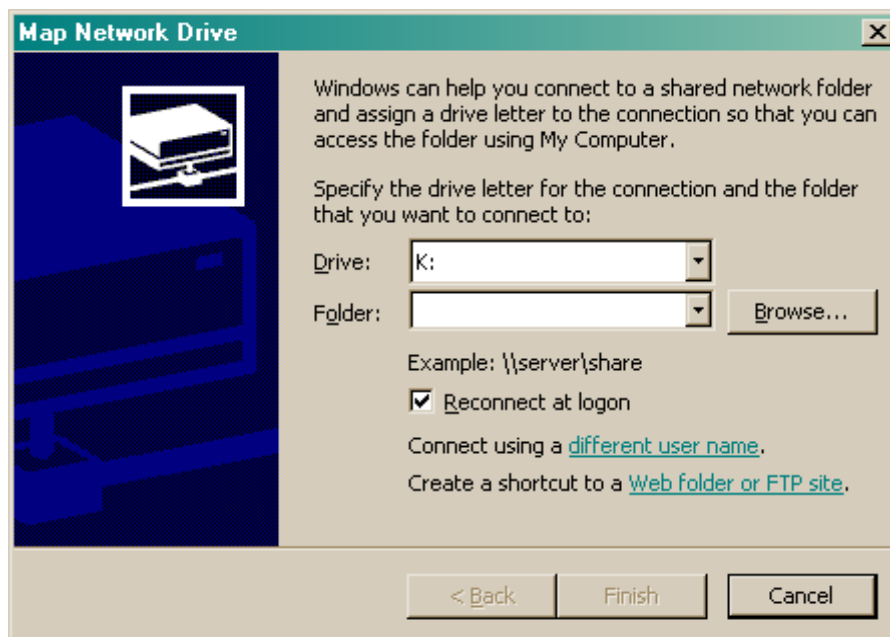
- It is necessary to log in twice to access the inbox. Once to the digital archive itself, and a second time to the actual inbox folder. The user name and password is the same for both log in. The user name is the set id. Both the set id and the password will be given to you by the transfer archivist.
- You must not create any sub-folders in the inbox folder. Any VEOs in the sub folder will be ignored during transfer.
- Once all the VEOs have been copied to the inbox, an end-of-set trigger file must be created in the inbox. This presence of this file is detected by the digital archive and invokes the processing of the set.

An illustration of the process of exporting to PROV over the Internet is shown below. The process manually copies the files to the PROV inbox using Windows Explorer to connect to the digital archive.

Prior to commencing the export, the transfer archivist will communicate the following information:

- The DNS address of the digital archive. This is 'inbox.prov.vic.gov.au'.
- Manifest used to create the Set: In this case it is: 7591p2u119batch112.xml
- Set Name: In this case it is: TR2005-0053-VPRS7591-P0018-S3380
- External Inbox Password. In this case it is 'whtdptgl'

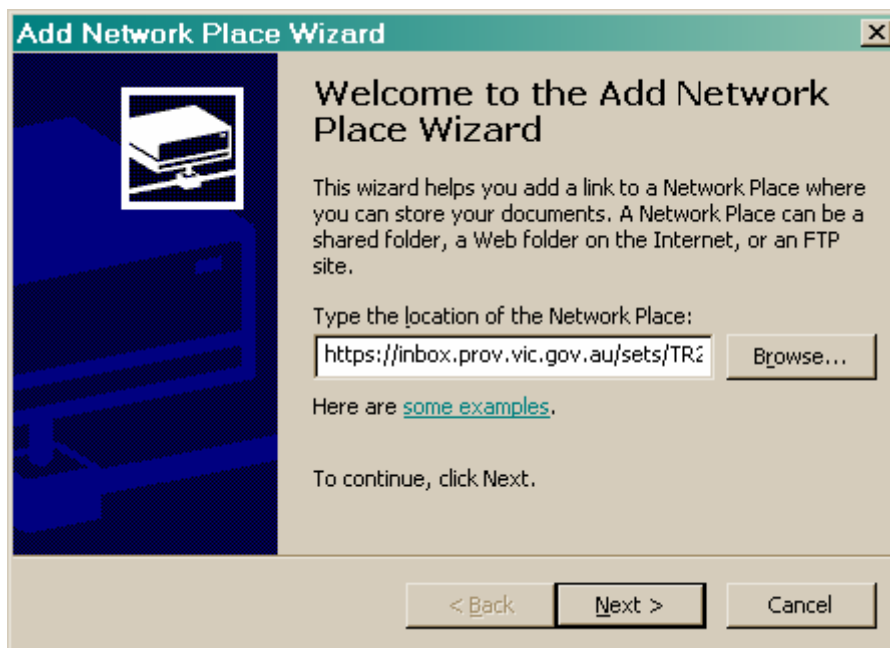
The first step in transferring involves opening a window on the inbox on the digital archive. From the main window of Windows Explorer, select 'Map Network Drive...' from the Tools menu of Windows Explorer.





Select the 'Create a shortcut to a Web folder or FTP site' option at the bottom of the screen and enter the URL of the external drop box that has been created for this set. The URL in this case is <http://inbox.prov.vic.gov.au/sets/TR2005-0035-VPRS7591-P0018-S3380>. This is composed of the:

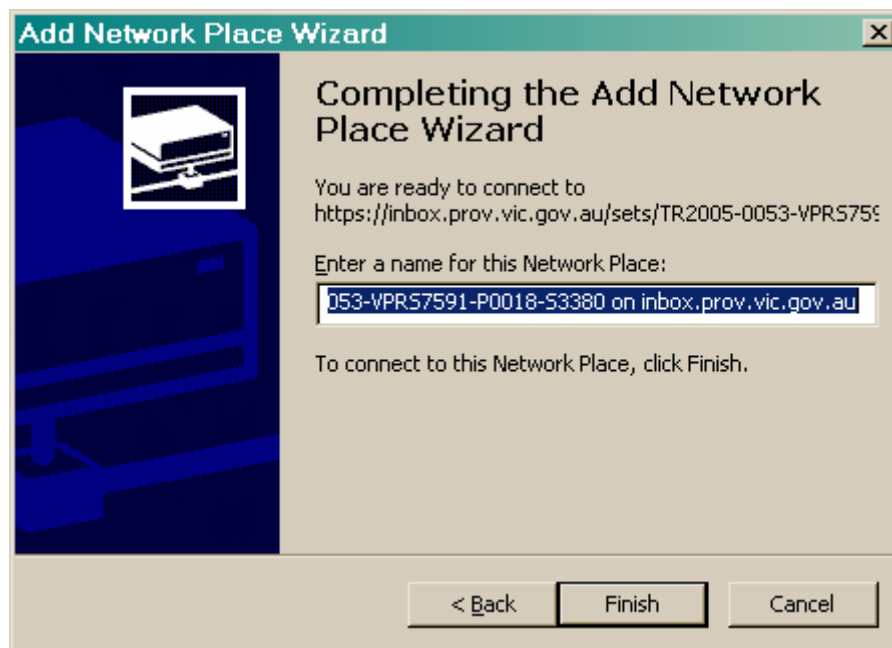
- protocol (HTTP)
- IP address of the digital archive.
- folder 'sets'
- Set id



Select 'Next'. Access to the digital archive is protected and requires a user name and password to enter the server. The user name is the name of the set, and the password will be given to you by the transfer archivist in the email.



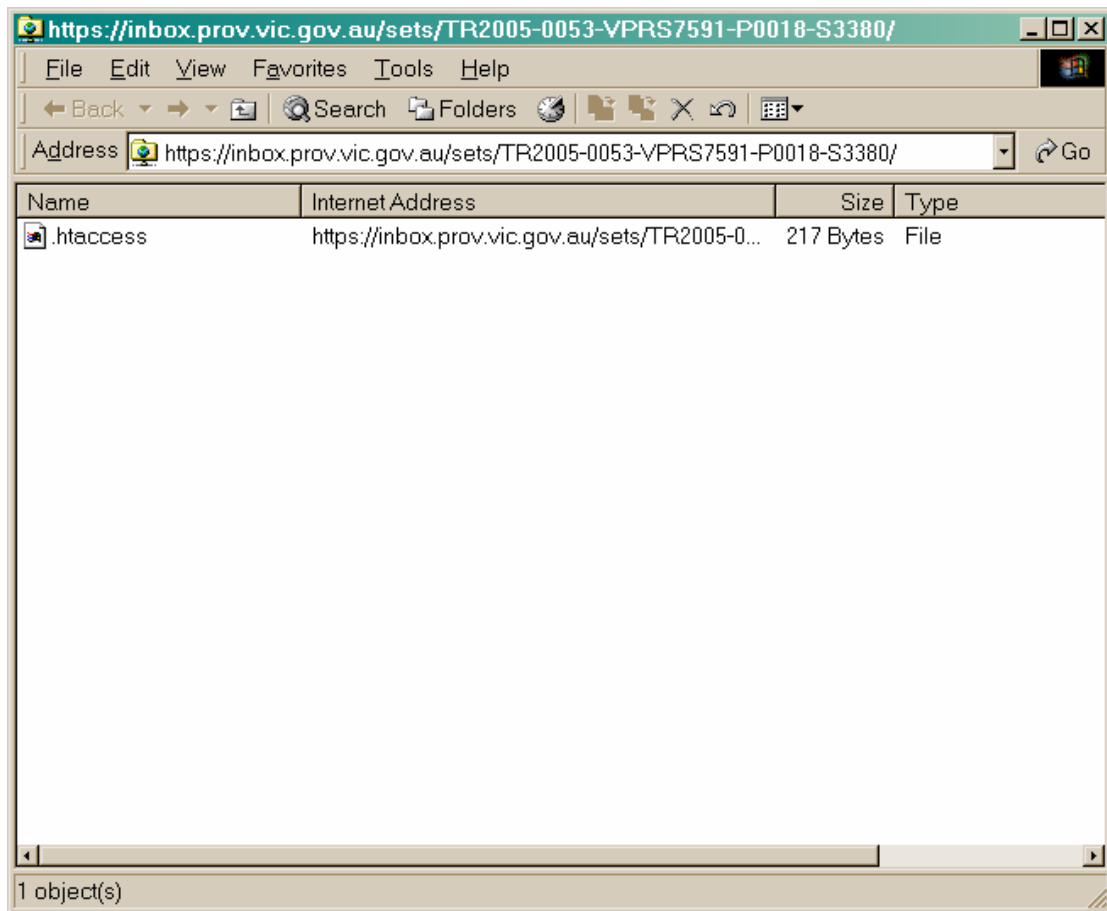
Select 'OK'. This will take you back to the 'Add Network Place Wizard' from which you must select 'Finish'.



The first login allowed access to the server holding the inbox. It is now necessary to log into the actual inbox itself. The username and password are the same as used when logging into the digital archive previously.

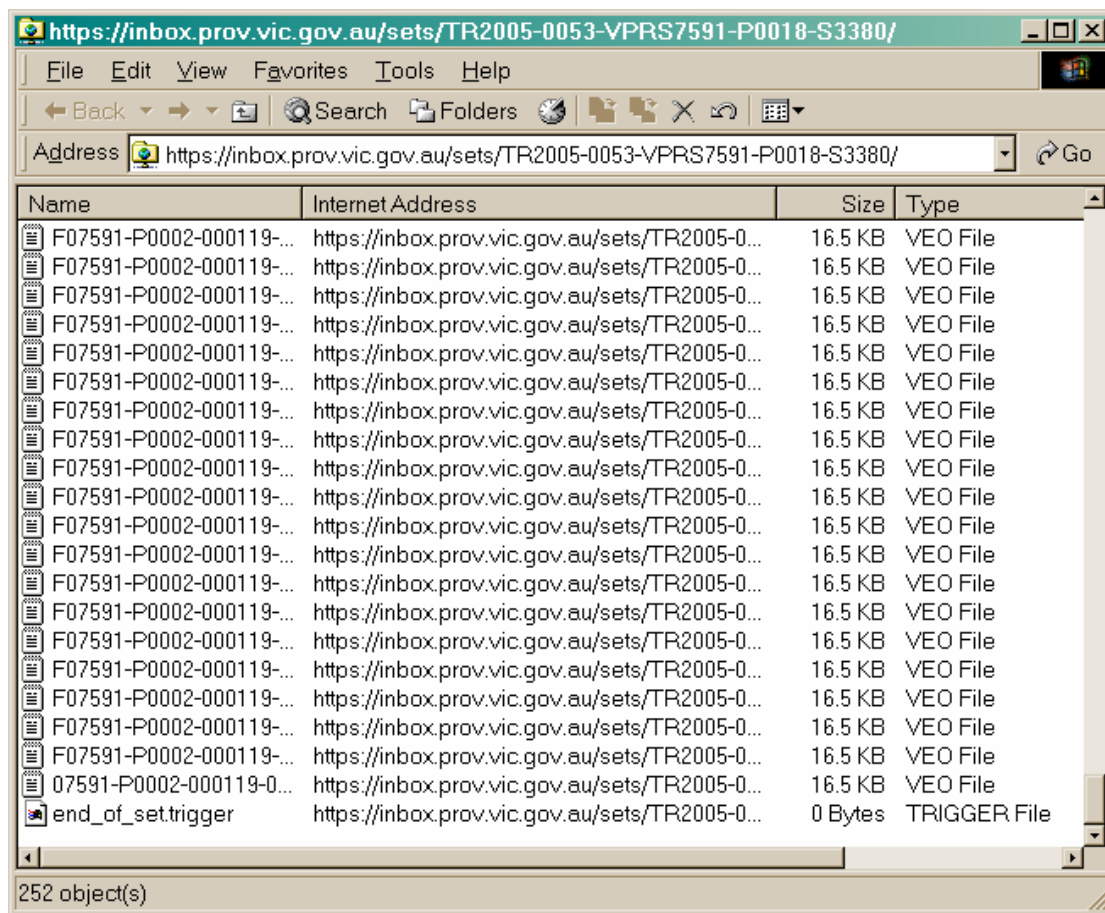


This will display the contents of the internal inbox for that set. This should be empty except for the '.htaccess' file.





Once the VEOs have been copied into the inbox, the end-of-set trigger file must be created in the inbox. The end-of-set trigger file is an empty (zero length) file.



Processing of the set will commence automatically once the digital archive detects the presence of the end-of-set trigger file in the inbox directory. The first step in processing involves moving the VEOs from the inbox into the digital archive itself. The VEOs will disappear from the inbox as they are moved into the digital archive.

## 8 Export on Physical Media

### 8.1 Transport of Media

Sensitive records should be protected against unauthorised access during transfer from the agencies to PROV.

As well, tape media should not be subject to large temperature or humidity fluctuations, nor excessive vibration, nor dirty environments. Suitable lockable carry cases are available for secure transport and protection from the environment.

When tape media arrive at a location, they should be kept in the machine room for at least 24 hours to allow the temperature to equilibrate before attempts to use the media.

## 8.2 Labelling of Media

Media should be labelled, both for humans and machines.

For humans, an external visible label identifies media and reduces mix-ups, including the accidental over-writing of data. The visible label can also form the external label for computers (see below), but it could include other information, e.g. dates, cycle numbers. It is useful for the export to be labelled with contact information to aid recovery in case of loss in transit.

For computers, labelling is of two forms – an external label and an internal label. The external label is often a barcode combined with OCR characters which robotic systems can read. The external label is used to ensure the correct media is loaded onto a drive. As well, for proper control, all media should be labelled internally, that is, should have a copy of the external label written onto the tape or CD itself.

Having correctly labelled tapes allows error checking, and also provides for multi-file volumes (allowing easy writing of multiple sets of data, possibly at different times) and provides for multi-volume files (allowing files larger than a volume to be handled on multiple tapes). Without the labelling control, this sort of capability would need to be handled with commands or scripts, which is not desirable, and is error-prone.

There are no standards for the internal labelling of CDs, so PROV requires the inclusion of a file named 'Label.txt' at the root level directory. The contents of this file must duplicate the information on the external label.

When using tape media the internal label is particularly important. Agencies should label tapes using the 'ANSI' tape label format [ISO1001] if the software/hardware being used to generate the tape supports this format.

The tape driving software should be aware of labelled tapes, and the conventions for labelling of volumes. It is important to realise that the initial internal labelling of a tape by a tape drive is an event which needs to be carefully managed, since the process writes on a tape without checking any pre-existing information on the tape, and has the potential to destroy the contents of media mistakenly labelled. (Tape media are sequential devices – anything written at the beginning of a tape renders any existing information on the tape unrecoverable by normal means, even if almost all of the data is not physically over-written.)

## 8.3 Export Media

When selecting the approved export media a major goal was to provide a limited range of media that would be suitable for large agencies transferring large numbers of records as well as small agencies transferring small numbers of records. Only a limited list of media types have been approved as each additional media type adds to PROV's costs.

In selecting export media, PROV has used the following criteria:

- *Media capacity.* A range of media capacity is supported to support large and small exports of records. There is a preference, however, to support large capacity media as this is usually cheaper and more readily available (it is hard to purchase and use older, lower capacity, media).
- *Cost.* There are two costs associated with each type of media: the cost of the media itself and the cost of the hardware used to read and write the media. Again, the goal has been to provide a range of cost options, from high capacity, expensive, media used for major systems to relatively low capacity, cheap media used by PC-type servers.

- *Ubiquity*. The ideal media is ubiquitous; readers and writers already exist on most systems and so there is no additional cost to generate the transfer media.

The following media types are approved for export to PROV:

- *CDs*. Although of relatively low capacity (650 MB), CD writers (burners) are very common and extremely cheap. The writeable media is cheap, and with suitable care in selection and production, of reasonable archivable suitability.
- *DVDs*. These are essentially a higher capacity (4.7 GB) replacement for CDs. Currently less common than CDs, the price and availability of media and writers should drop rapidly over the next year or so as they become common consumer items. Archival suitability is slightly questionable, but should be similar to CDs.
- *DDS*. This tape format is extremely common as it has reasonable capacity. Both the media and hardware are very cheap.
- *Ultruim LTO*. This tape format is expected to become popular over the next year or so as it is supported by several independent vendors and has a very high capacity.

### 8.3.1 CDs

Producing compact discs is easy; producing high quality recordable compact discs that will survive several years of storage is complex. The various restrictions on the use of CDs as export media are intended to ensure high quality CDs that will survive several years of storage at PROV. Media Sciences has an excellent site on the issues surrounding the production of high quality compact discs (<http://www.mscience.com/> visited 30 June 2006) and many of the restrictions PROV imposes are derived from information on this site. Another useful Web site is the CD-Recordable FAQ (<http://www.cdrfaq.org/> visited 30 June 2006).

Media Sciences states that the best way to ensure longevity of a CD is to start with a high quality recording as it is difficult to determine the quality of a blank CD. It is important to realise that burning an 'error free' disc is not the end of the story. An apparently error-free disc almost certainly will contain errors, but these will be masked by the error correction added when the CD is burnt. The number of errors will depend on:

- the original quality of the disc. The quality of the disc varies dramatically between vendors, between vendor production facilities, and between production batches.
- interactions between the disc and the writer. Many factors can result in the CD writer producing a CD with a high error rate. For example, the laser in the CD loses power over time, which can result in less accurate CD burning. A more complex issue is the different types of dyes used in recordable CDs, which require different use of the laser. This means, for example, that a CD writer optimised for one type of dye may do a poorer job with a disc that uses a different dye.

The underlying error rate would be of less importance if it did not increase over time as a CD deteriorates. Chemical deterioration is of less significance than mechanical deterioration. It is not widely realised how fragile the label side of a CD is. The label is very thin and every scratch adds errors to the CD and these cannot be removed.

The result of all this is that a CD with a high underlying error rate may appear to be perfect when first produced, but become unreadable after a short period of use. A CD with a low underlying error rate has a greater buffer against the unavoidable damage incurred during its life.

The reasons for some of the restrictions and qualifications in using CDs are discussed below:

- *CD Capacity.* Higher capacity CDs are produced by tightening the hardware tolerances, for example, using the minimum possible track spacing. This means that data is written closer together and this can result in higher error rates, or some CD players not reading the disc. Media capacities over 80 minutes are outside the CD standard and must never be used for burning CDs for PROV.
- *Manufacturer.* CD quality varies enormously; and thus we would strongly recommend buying quality media. Subject 7-4-1 of the CDR FAQ (<http://www.cdrfaq.org/> visited 30 June 2006) comments on the reputation of various manufacturers.
- *File System Formats.* Not all CD readers or operating systems support all file system formats. These formats are those which PROV can use. The ISO 9660 file system format only supports filenames similar to the archaic MS-DOS format: eight characters and a three character extension, uppercase letters and numbers, and '.' and '\_'. With modern computers, support for longer filenames is essential and so PROV will accept the Joliet file system. Rock Ridge is designed for UNIX-based systems and HFS for Macintosh systems; these are not acceptable, as PROV systems are currently based on PC technology. ISO 9660:1999 is a newer, and more powerful, version of ISO 9660. Not all operating systems support ISO 9660:1999.
- *Labelling.* All pieces of media must be clearly labelled. The trick is to ensure that the label does not damage the label surface of the CD which, as noted before, is extremely fragile. Ball-point pens will almost certainly scratch the label side of a CD, so felt-tipped pens are the only types accepted. The use of solvent-based felt-tipped pens (e.g. permanent markers) may result in chemical reactions with the aluminium label surface and so are prohibited. Adhesive labels may detach from the CD, or the adhesive may cause chemical reactions with the surface of the CD. Poor application of the label may even pull the aluminium surface of the CD off. Off-centre application of a label may unbalance the CD, making it difficult or impossible to achieve CD balance when it is spinning in a reader.
- *Packaging.* The major cause of deterioration of CDs is scratches on the label side. Requiring discs to be contained in jewel cases means that they can be physically protected against damage both during export and during storage. 'Jewel' cases are the plastic cases similar to those that pre-recorded CDs are sold in.

### 8.3.2 DVDs

From a technology perspective, DVDs are almost identical to CDs and many of the comments applying to the longevity of CDs also apply to DVDs; particularly those concerned with the fragility of discs.

One positive feature of DVDs, compared with CDs, is that there are fewer issues with recording modes and file system formats. This is because DVDs are a newer, more sophisticated, design, and, perhaps, because there has yet to be sufficient time for incompatible extensions to accrete to the standard.

A problem with DVDs is the number of incompatible recordable formats. These compatibility problems are largely due to the different recording materials. These have different reflectivity to the read lasers and, hence, one recordable technology may not read another. CDs have similar problems, but the issue is more acute with DVDs due to the higher densities and tighter tolerances.

For this reason, we are requiring DVD-R or DVD+R (i.e. recordable) DVDs rather than the rewritable or erasable DVDs. Recordable DVDs use similar technology to recordable CDs and currently seem to have the least problems with incompatibility between readers.



For simplicity we require single-sided DVDs (it is possible to obtain double-sided recordable DVDs, but these are uncommon).

### 8.3.3 Tape Media

Magnetic tape drives have been used in the computer industry for over 50 years and provided the first mass storage.

The enduring nature of magnetic tape is illustrated by the fact that reel-to-reel ('round') tapes are still used, at least in Australia, for the interchange of financial data, e.g. for payrolls between organisations and banks. Reel-to-reel tape has essentially been unchanged for about 40 years, with the format on the media (9-track and 6250 bits per inch) unchanged for about 20 years.

The enduring nature of reel-to-reel tape is unlikely to be repeated for more modern tape formats, as there were many manufacturers of both the reel-to-reel drives and the media. Now, except for developments in the last three years, each drive type is made by only a few vendors (in several cases only one), and the media is similarly restricted to few vendors. This leads to a rapid turnover of drives and types. Further disk drive technology is challenging tape for capacity per dollar. Indeed, it is only possible to justify tape for static storage when very large volumes need to be managed. However, tape remains useful for the interchange of moderate to large volumes of data, as disks are traditionally not transportable, and fast networks are not yet sufficiently ubiquitous.

So the challenge is to identify a suitable magnetic media technology which meets the following criteria, listing here in approximate descending order of importance for this target application:

- market penetration
- number of vendors
- interchangeability
- already in use in Victorian Government
- server compatibility
- durability
- media robustness
- media shelf life
- cost – drive and media
- capacity
- speed
- physical size – drive and storage space
- automation capability (tape library)
- access time.

An excellent overview of tape technologies has been produced by Gartner Research and has been made available by the Tape Council [GART1]. The following table gives some of the characteristics of tape types.

Media type	Drive manufacturer	Tape technology (L - linear	Tape width max.	Capacity gigabyte uncom-	Read/ write speed megabyte

		H - helical)	(mm)	ressed	per second
SLR 100	Tandberg	L	?	50	5
Travan	Seagate Imation	L	?	20	2
DLT 4000	Quantum Tandberg	L	12.7	20	1.5
DLT 7000	Quantum Tandberg	L	12.7	35	5
DLT 8000	Quantum Tandberg	L	12.7	40	6
DLT1	Benchmark	L	12.7	40	3
SuperDLT (SDLT)	Quantum	L	12.7	110	11
SuperDLT 2	Quantum	L	12.7	160	16
Ultrium LTO	HP, IBM, Certance	L	12.7	100	20
Ultrium LTO 2	HP, IBM, Certance	L	12.7	200	40
3480	IBM	L	12.7	1	3
3490E	IBM	L	12.7	0.4 or 0.8	9
3590/3590E/ 3590H	L	L	12.7	60	14 to 100
9840	StorageTek	L	12.7	20	19 to 70
9940	StorageTek	L	12.7	200	30 to 70
DDS-2	several	H	4	4	
DDS-3	several	H	4	12	1
DDS-4	several	H	4	20	3
Mammoth-2	Exabyte	H	8	60	12
VXA-2	Exabyte	H	8	80?	12
AIT 2	Sony	H	8	50	6
AIT 3	Sony	H	8	100	12

Given the characteristics, the best tape transfer options are:

- *Ultrium LTO*. Already in use by at least two agencies. There are multiple vendors of both tape and drive. Large capacity (100 gigabytes). Cheap media (per gigabyte).
- *DDS*. DDS tapes are already in use by at least two agencies, and there is a very large installed base. The tape drives and media are very cheap and suitable for connection to smaller systems (e.g. PC-based servers). There is a concern with interchangeability of media, but this will be handled by reading each piece of media upon receipt by PROV.

## 8.4 Media Management Software

It is necessary to have a variety of software in order to transfer computer files to and from media. This software includes:

- volume management software
- media mounting software

- media driver software
- data writing software.

Since tape use at many sites is usually confined to backups, often totally under the control of a canned backup application, there is often little expertise in tape management and use. A simple solution for production of media for export to PROV is needed, but this may be difficult to achieve.

#### 8.4.1 Volume management

Volume management software is effectively a database system that maintains a register of media. This register flags the media status, capacity, time of first and last use, amount of data written, number of mounts, etc.

It is likely that only PROV will need to use this software.

#### 8.4.2 Media mounting software

Software needs to be available to allow the mounting of media volumes. This software will need to be available at both the agency and PROV, although it need not be the same software.

With CDs this is normally part of the operating system and is carried out automatically.

Tape, however, is more complex, and the functions supported (e.g. multi-volume files) are more sophisticated. Tape mounting is usually carried out through a library system (e.g. StorageTek's ACSLS). The host component of this software needs to monitor the status of read and write controls. This disallows, for example, the mounting of a tape for writing when only the read access is flagged.

Tape mounting software needs to support the handling of multi-file tapes, multi-volume files, tape labelling, and tape positioning (especially for appending).

#### 8.4.3 Media driver software

Software needs to be available to allow driving of the read/write hardware. This software is normally provided by the driver vendor, and it can be difficult to obtain for unusual host systems. Lists of supported devices/operating systems are usually available from computer or drive vendors.

#### 8.4.4 Data writing software

Once the tape mounting and driving software has done its job, software is needed to actually control the copying of data from disk files to the media and vice versa. Matching software needs to be available at the agencies and PROV.

Since tape does not usually have a file system, software is required to allow the archiving of multiple files into a single stream which is then written to the tape.

In UNIX, this has traditionally been done with utilities like tar ('tape archive') and cpio. However, since 1990 there has been a standard POSIX-compliant utility called pax ('Portable archive interchange') [PAX]. Implementations of Pax are available on UNIX-based systems (Solaris, UNICOS and Linux) and also under the various Microsoft operating systems (e.g. under Interix, available in Microsoft Windows Services for UNIX 3.0).

There does not appear to be an equivalent standardised software for PC based systems.

Implementations of these utilities need to be checked to see that they do not have limits imposed such as 2 gigabyte archive limits.

The other option for data writing software would be to use one of the backup utilities. PROV decided against this option as the formats used are specific to the particular vendor and, often, to the particular version of the software. Unless it is possible to standardise backup utilities across the Victorian Government, accepting tapes written using a vendor's backup format was considered too costly for PROV. Further, it was felt unlikely that it would be possible to read these proprietary formats in the future.

#### 8.4.5 Summary of software requirements

In summary, the following software will be required:

- *Media driver software.* Required by both PROV and the agencies, this will be specific to the particular drive and host computer system.
- *Media mounting software.* Required by both PROV and the agencies, this will be specific to the particular host computer system.
- *Data writing software.* Required by both PROV and the agencies. PROV requires agencies to use PAX or TAR to write the tapes.
- *Volume management software.* Some form of volume management will be required by PROV, but not by the agencies.

## 9 References

- [GART1] Tape Drives: Overview, Adams & Adams, Gartner Research, 17 July 2002, <http://www.tapecouncil.org/pdf/Tape%20Drives%20Overview.pdf> visited 30 June 2006.
- [ISO1001] Information processing; File structure and labelling of magnetic tapes for information interchange, ISO 1001:1986.
- [PAX] Archive/Interchange File Format specified in ISO/IEC 9945-1:1990 (IEEE Std 1003.1-1990), Information technology - Portable Operating System Interface (POSIX) - Part 1: System Application Program Interface (API) [C Language].
- [WEBDAV] Extensions for Distributed Authoring – WEBDAV, Golan et al, IETF RFC 2518, February 1999, <http://www.ietf.org/rfc/rfc2518.txt> visited 20 April 2006.