

# Specification

---

**PROS 15/03 S3**  
**Long term preservation formats**

Version number: 1.0  
Issue Date: 1 June 2015  
Expiry Date: 1 June 2020

## Copyright Statement

© State of Victoria 2015



Except for any logos, emblems, and trade marks, this work (PROS 15/03 S3 Long term preservation formats) is licensed under a Creative Commons Attribution 3.0 Australia license, to the extent that it is protected by copyright. Authorship of this work must be attributed to the Public Record Office Victoria. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/au/>

## Disclaimer

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 Specification.

## Acknowledgements

The Public Record Office Victoria would like to acknowledge and thank the valuable contributions of people who took the time to comment during the development of this Specification.

Under section 12 of the *Public Records Act 1973*, the Keeper of Public Records is responsible for the establishment of Standards for the efficient management of public records and for assisting public offices to apply those Standards to records under their control. Officers in charge of public offices are responsible under section 13 of the Act for carrying out, with the advice and assistance of the Keeper, a program of records management in accordance with the Standards established under section 12 of the Act

# Table of Contents

---

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Relation to previous Standard	7
1.2	Related documents	7
1.3	General References	7
1.4	Acknowledgements	7
<b>2</b>	<b>VERS Long Term Preservation Formats</b>	<b>8</b>
2.1	How to choose one of the long term preservation formats	8
2.2	Long term preservation formats	9
2.3	Other formats	10
<b>3</b>	<b>References</b>	<b>12</b>

This page is blank

# Executive Summary

---

This specification is part of the *Standard for the encapsulation of digital information* (PROS 15/03). It describes the long term preservation formats that are acceptable to PROV.

Obsolescence of digital formats is a problem for long term digital preservation. When a format becomes obsolete it may be difficult (or impossible) to obtain software to read the format and render the information contained within it. A long term preservation format is one that is

- So ubiquitous that software to read the format is likely to be available for the foreseeable future, or
- Well documented so that readers can, if necessary, be implemented from scratch.

This page is blank

# 1 Introduction

This specification lists a set of long term preservation formats that are considered to have a long usable life, and that will minimise the cost of providing access to the content over its life.

All content transferred to PROV must be migrated (if necessary) to one of these formats. If content is migrated, the original format must also be transferred to PROV (unless specifically exempted).

It is recommended that

- Agencies adopt these formats when creating content (as this minimises subsequent migration costs).
- Agencies adopt these long term formats for content that they need to keep for a long time.

The formats were chosen on the following criteria:

- Extremely widespread adoption.
- Being the dominant format in a particular category.
- Multiple independent implementations of the software.
- A published formal specification that implementations adhere to.
- Already accepted by the previous version of this standard.

## 1.1 Relation to previous Standard

**In order to protect the investment already made by vendors and agencies, PROV will continue to accept VEOs conformant to the previous VERS Standard (PROS 99/007 (Version 2)) for the indefinite future.**

The new format, however, is not backwards compatible with the previous version.

## 1.2 Related documents

This document should be read in conjunction with

- PROS 15/03 Standard for the encapsulation of digital information
- PROS 15/03 Specification 1: Constructing VERS Encapsulated Objects
- PROS 15/03 Specification 2: Adding metadata to VEOs

## 1.3 General References

References to format specifications are included in the body of this specification.

## 1.4 Acknowledgements

We would like to acknowledge and thank the people who took the time to comment on earlier drafts of this proposal. Nearly all the comments have been included in this draft, and where this is not possible we have included footnotes to explain the reasons.

# 2 VERS Long Term Preservation Formats

---

This specification identifies the formats that are considered low risk for the long term preservation of information.

The risk being addressed is that in the future it will not be possible to obtain software to extract or present the information embedded in a digital object. Over a sufficiently long period of time all formats can be expected to become unreadable. At some point, then, it will be necessary to undertake a preservation action for a format. The goal of this specification is to identify formats for which this preservation action is likely to be a long time in the future, and the necessary tools are available when it is necessary to perform this preservation action.

## 2.1 How to choose one of the long term preservation formats

### 2.1.1 Requirement to use one of the formats

All record content transferred to PROV must be represented in one of the long term preservation formats in this Standard.

Agencies are strongly encouraged to use these long term preservation formats for information that must be retained for more than seven years, even if the information does not have to be transferred to PROV. Use of the formats will aid in ensuring long term access to the content. The longer the information needs to be kept, the more important it is to use one of these long term preservation formations.

### 2.1.2 Avoiding migration

Agencies are strongly encouraged to adopt these long term preservation formats for day to day business use. This will avoid the requirement to subsequently migrate information to a long term preservation format.

Migration from one format to another is to be avoided. In general, migration is expensive (to obtain the necessary tools, to carry out the migration, and to conduct the necessary quality assurance to ensure the migration was carried out successfully). There is always a risk of losing information when migrating.

### 2.1.3 Format selection

Several long term preservation formats are provided for most categories of information.

Agencies and agency staff can choose the most appropriate format for their business needs. In accordance with our policy of avoiding migration, the most appropriate format to choose will normally be the one in which the business is actually undertaken.



## 2.1.4 Version selection

Unless otherwise indicated, PROV will accept any version or variant of the selected long term preservation formats. This is because, in most cases, it is difficult for agencies to configure or set up software products to produce particular variants of a format. Further, most software will produce the latest version<sup>1</sup> of a particular format.

Where particular versions of a format are preferred over others this is indicated. Agencies are encouraged to adopt the recommended versions where possible.

## 2.1.5 Files must be valid according to the format specifications

PROV reserves the right to reject records that are not valid according to the format specification.

## 2.1.6 Include the original format

Where record content has been migrated to a long term preservation format for the purposes of transfer, a copy of the original, un-migrated, format must also be included in the VEO unless otherwise agreed by PROV.

The requirement to include a copy of the original format guards against the following risks:

- that the migration did not result in an accurate representation of the original record
- that the migration caused a loss of information
- that a better migration approach may be available in the future

PROV will not generally require a copy in original format where

- The record is extremely large, and
- The migration process is a routine technology process with little chance of content loss

A typical example of a situation where the original would not normally be required is the conversion of video or audio to a long term preservation formation.

## 2.2 Long term preservation formats

The following formats are accepted by PROV as long term preservation formats:

- Document and text formats:
  - Plain text (.txt)
  - Portable Document Format (.pdf). It is preferred that PDF documents be conformant to PDF/A-1 (ISO 19005-1) or PDF/A-2 (ISO 19005-2). PDF/A-3 must not be used.
  - Microsoft Word (.doc, .docx)
- Web formats:
  - HTML (.htm, .html, .css). It is recommended that HTML files conform to HTML 4.01 standard (<http://www.w3.org/TR/1999/REC-html401-19991224/>) and CSS 2.1 (<http://www.w3.org/TR/2011/REC-CSS2-20110607/>)
  - eXtensible Mark-up language (.xml)<sup>2</sup>
  - Web ARChive format (.warc)<sup>3</sup>

---

<sup>1</sup> At the time the creation software was produced.

<sup>2</sup> XML files will be readable for the indefinite future, but they may not be interpretable. This is because the meaning of the XML markup is defined in separate standards (e.g. SVG for vector graphics).

<sup>3</sup> Note that WARC is a container format that encapsulates web objects. Each web object (e.g. web pages) in the WARC file must be in one of the long term preservation formats specified in this specification.

- Spreadsheet formats:
  - Comma separated values (.csv)<sup>4</sup>;
  - Microsoft Excel (.xls, .xlsx)
- Presentation formats:
  - Microsoft PowerPoint (.ppt, .pptx)
  - Portable Document Format (.pdf). It is preferred that PDF documents be conformant to PDF/A-1 (ISO 19005-1) or PDF/A-2 (ISO 19005-2). PDF/A-3 must not be used.
- Image Formats:
  - JPEG (.jpg, jpeg);
  - JPEG2000<sup>5</sup> (.jp2);
  - Tagged image file format (.tif, .tiff)
- Audio Formats:
  - MPEG 1/2 Audio Layer 3 (.mp3);
  - MPEG-4 (.mp4)
  - WAVE (.wav) using an LPCM codec
- Video Formats:
  - MPEG-4 (.mp4)
- Email Formats:
  - MIME (.eml).

## 2.3 Other formats

Agencies are encouraged to contact PROV if a common format they use is not in this list<sup>6</sup>. However, before extending this list of formats PROV will work with the agency to determine the most appropriate long term preservation format.

Selection of long term preservation formats is based on the assumption that ultimately any format is likely to fall out of use and objects in that format will require preservation actions.

Good long term preservation formats are those that are likely to have a long lifespan before preservation interventions are required, and that when a preservation action is required suitable tools should be easily obtained. It is important to note that this Standard does not assume that formats will have an indefinite lifespan.

Characteristics that suggest that a format is likely to have a long lifespan before preservation interventions are required are that:

- The format is in extremely widespread use
- The format has the dominant market share in its domain.

These two characteristics mean that the market is expected to sustain the format. New products in the domain must accurately support the format (otherwise it is extremely difficult for them to gain market share). The number of objects in that format means that there is an economic incentive for developers to produce readers or migration tools for that format, even if the original vendor ceases support. An additional benefit of selecting common, dominant, formats is that these are likely to be the majority of record content held by an agency.

---

<sup>4</sup> See <http://tools.ietf.org/html/rfc4180> for a non-normative definition of a CSV file.

<sup>5</sup> JPEG2000 is accepted as a long term preservation format in PROS 99/007 (Version 2.0). For this reason, PROV will continue to accept files in this format, however, at the moment (2015) it would fail the economic adoption test.

<sup>6</sup> This applies, in particular, to appropriate long term preservation formats for specialized types of data (e.g. CAD, GIS).

Characteristics that suggest that tools will be available to undertake the preservation action include:

- A published format specification exists<sup>7</sup>.
- Multiple independent implementations exist of format creators/readers<sup>8</sup>.

---

<sup>7</sup> It is not necessary that the format be published by a standards body, but this is preferred.

<sup>8</sup> Multiple independent implementations show 1) that there is an economic incentive to read the selected format, and 2) that the format can be accurately implemented by independent vendors.

# 3 References

---

## Other Resources

You can obtain relevant publications, supplies of relevant forms, and answers to any enquiries you may have by first contacting your agency's records manager or the Public Record Office Victoria:

For more information on digital preservation, please contact:

Government Services  
Public Record Office Victoria  
Ph: (03) 9348 5600  
Fax: (03) 9348 5656  
Email: [agency.queries@prov.vic.gov.au](mailto:agency.queries@prov.vic.gov.au)  
Web: [www.prov.vic.gov.au](http://www.prov.vic.gov.au)