

# Public Record Office Victoria

## SPECIFICATION

### PROS 19/05 S3: LONG TERM SUSTAINABLE FORMATS

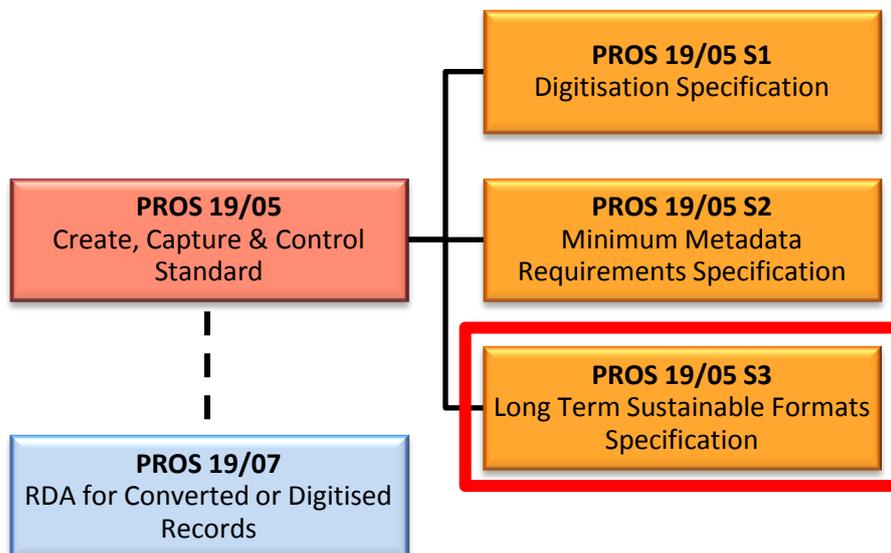
Version number: 1.0  
Issue Date: 16 August 2019  
Expiry Date: 16 August 2029

#### About this Specification

All Victorian public offices must comply with this Specification when preparing digital records for transfer to PROV. Permanent value digital records must be in one of the specified formats or able to be converted reliably and efficiently to one of these formats.

It is recommended that Victorian public offices also use these formats when temporary value records need to be kept for a long period of time.

The diagram below shows the relationship between this Specification and related documents.



# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Authority of Standards and Specifications	3
1.2	Obligation	3
1.3	Applying this Specification	3
<b>2</b>	<b>Format Selection</b>	<b>4</b>
2.1	Criteria for Format Selection	4
2.2	Format Requirements	4
2.3	Avoiding Migration	4
2.4	Format Selection	4
2.5	Version Selection	5
2.4	File Validity	5
2.5	Inclusion of Original Format	5
<b>3</b>	<b>Long Term Sustainable Formats</b>	<b>6</b>
<b>4</b>	<b>Other Formats</b>	<b>7</b>

# 1 Introduction

## 1.1 Authority of Standards and Specifications

Under section 12 of the *Public Records Act 1973*, the Keeper of Public Records ('the Keeper') is responsible for the establishment of Standards for the efficient management of public records and for assisting Victorian public offices to apply those Standards to records under their control.

Heads of public offices are responsible under section 13b of the *Public Records Act 1973* for carrying out a program of efficient management of public records. The program of records management needs to cover all records created by the public office, in all formats, media and systems across the organisation.

The Standards and Specifications support the Victorian Electronic Record Strategy (VERS) Digital Forever 2018-2021<sup>1</sup>, which is designed to ensure the creation, capture and preservation of authentic, complete and meaningful digital records.

This Specification is part of the *PROS19/05 Create, Capture and Control Standard*. This Specification, as varied or amended from time to time, shall have effect for a period of ten (10) years from the date of issue unless revoked prior to that date.

## 1.2 Obligation

It is mandatory for all Victorian public offices to follow the principles and comply with the requirements of the Standards and Specifications issued by the Keeper.

## 1.3 Applying this Specification

This Specification sets out the formats that have been assessed by the Public Record Office Victoria (PROV) as likely to remain readable and usable for a long period of time and provides guidance on format selection.

Permanent value digital records must be in one of these formats or able to be converted reliably and efficiently to one of these formats. This will help to ensure:

- records are readable and usable while held by the public office; and
- VERS encapsulated objects (VEOs) can be constructed when records are being prepared for transfer to PROV.<sup>2</sup>

It is recommended that these formats also be used when temporary value records need to be kept for a long period of time.<sup>3</sup> This will help to ensure records are readable and usable for the time they need to be retained and records can reliably and effectively be migrated during system or organisational changes (e.g. machinery of government changes, administrative mergers etc.).

---

<sup>1</sup> The previous *PROS15/03 Standard for the encapsulation of digital records* has been revoked and the requirements have now been included in the *PROS19/05 Create, Capture and Control Standard* and associated Specifications.

<sup>2</sup> If permanent value digital records planned for transfer are not in one of these formats or can't be reliably and efficiently converted to one of these formats, seek advice from PROV.

<sup>3</sup> For example, if the records must be retained for longer than the life of the system. As it is not necessarily known when formats are being determined, an arbitrary time period might be selected. For example, it might be decided that temporary records which must be retained for longer than 10 years will be held in one of the formats set out in this Specification.

# 2 Format Selection

## 2.1 Criteria for Format Selection

There is a risk that in future it will not be possible to obtain software to extract or present information embedded in a digital object. Over a sufficiently long period of time, all formats can be expected to become unreadable. At some point it will be necessary to undertake a preservation action for a format. This specification aims to identify formats which are likely to be available for a long period of time and for which the necessary tools are likely to be available when records need to be exported/migrated in the future.

The formats were chosen using the following criteria:

- extremely widespread adoption
- being the dominant format in a particular category
- multiple independent implementations of the software or
- accepted by a published formal specification.

## 2.2 Format Requirements

All record content transferred to PROV must be represented in one of the long term sustainable formats outlined.

Public offices are strongly encouraged to also use these formats for digital records that must be retained for a long period of time. 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 sustainable formats.

## 2.3 Avoiding Migration

Public offices are strongly encouraged to adopt these long term sustainable formats for day to day business use. This will avoid the requirement to subsequently convert records to these formats.

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.

## 2.4 Format Selection

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

Public offices can choose the most appropriate format for their business needs. In accordance with avoiding migration, the most appropriate format to choose will normally be the one in which the business is actually undertaken.

## 2.5 Version Selection

Unless otherwise indicated, PROV will accept any version or variant of the selected long term sustainable formats. This is because, in most cases, it is difficult for public offices to configure or set up software products to produce particular variants of a format. Further, most software will produce the latest version<sup>4</sup> of a particular format.

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

## 2.4 File Validity

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

## 2.5 Inclusion of 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
- the migration caused a loss of information
- 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 format.

---

<sup>4</sup> At the time the creation software was produced.

# 3 Long Term Sustainable Formats

The following formats have been assessed by PROV as being long term sustainable:

Table 1: Sustainable formats.

TYPE	SUSTAINABLE FORMAT
Document and text	<ul style="list-style-type: none"> <li>Plain text (.txt)</li> <li>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). <b>PDF/A-3 must not be used.</b></li> <li>Microsoft Word (.doc, .docx)</li> </ul>
Web	<ul style="list-style-type: none"> <li>HTML (.htm, .html, .css). It is recommended that HTML files conform to <a href="#">HTML 4.01 Specification</a> and <a href="#">CSS 2.1 Specification</a>.</li> <li>eXtensible Mark-up language (.xml)<sup>5</sup></li> <li>Web ARChive format (.warc)<sup>6</sup></li> </ul>
Spreadsheet	<ul style="list-style-type: none"> <li>Comma separated values (.csv)<sup>7</sup></li> <li>Microsoft Excel (.xls, .xlsx)</li> </ul>
Presentation	<ul style="list-style-type: none"> <li>Microsoft PowerPoint (.ppt, .pptx)</li> <li>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). <b>PDF/A-3 must not be used.</b></li> </ul>
Image	<ul style="list-style-type: none"> <li>JPEG (.jpg, jpeg)</li> <li>JPEG2000<sup>8</sup> (.jp2)</li> <li>Tagged image file format (.tif, .tiff)</li> </ul>
Audio	<ul style="list-style-type: none"> <li>MPEG 1/2 Audio Layer 3 (.mp3)</li> <li>MPEG-4 (.mp4)</li> <li>WAVE (.wav) using an LPCM codec</li> </ul>
Video	<ul style="list-style-type: none"> <li>MPEG-4 (.mp4)</li> </ul>
Email	<ul style="list-style-type: none"> <li>MIME (.eml)</li> </ul>
Relational database	<ul style="list-style-type: none"> <li>SIARD (.siard)<sup>9</sup></li> </ul>

<sup>5</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>6</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.

<sup>7</sup> See <http://tools.ietf.org/html/rfc4180> for a non-normative definition of a CSV file.

<sup>8</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.

<sup>9</sup> SIARD format description is available here: <https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>.

# 4 Other Formats

Public offices are encouraged to contact PROV if a common format being used is not in the aforementioned list<sup>10</sup>. However, before extending this list of formats, PROV will work with the public office 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 suitable tools easily obtainable when a preservation action is required.

It is important to note that this specification does not assume that formats will have an indefinite lifespan.

Characteristics which 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 and 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 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 records content held by a public office.

Characteristics that suggest that tools will be available to undertake the preservation action is based on the knowledge that a published format specification exists<sup>11</sup> and multiple independent implementations exist of format creators/readers<sup>12</sup>.

## Copyright Statement

© State of Victoria 2019



Except for any logos, emblems, and trade marks, this work is licensed under a Creative Commons Attribution 4.0 International 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 <https://creativecommons.org/licenses/by/4.0/legalcode>

## 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 Standard.

---

<sup>10</sup> This applies, in particular, to appropriate long term preservation formats for specialised types of data (e.g. CAD, GIS).

<sup>11</sup> It is not necessary that the format be published by a standards body, but this is preferred.

<sup>12</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.