

Public Record Office Victoria

Case Study

VEO Vexation

A process to repair and re-sign approximately 56,000 error VERS Encapsulated Objects (VEOs)



Public Record
Office Victoria



Public Record Office Victoria Case Study 2015

VEO Vexation



Digital records are transferred to Public Record Office Victoria (PROV) in the form of VERS Encapsulated Objects, or VEOs. Prior to storage in the PROV Digital Archive, VEOs undergo a variety of checks ranging from virus scanning to digital signature validation. VEOs that fail validation are automatically rejected from the ingest process into PROV's Digital archive.

Key Topics

Digital archiving, VERS Encapsulated Object (VEO), stream editing, digital signature.

Key Discoveries

- **High volumes of VEOs with errors can often be repaired with relatively minimal effort with the use of online tools and commands.**
- **Having consistent formatting in all fields of metadata will help ensure effective VEO creation.**

The Challenge

Recently, some 56,000 VEOs awaiting ingest into PROV's Digital Archive failed validation. The reason was simple, but the consequences were quite severe. The VEOs were invalidated simply because they were found to have an additional colon and two zeros in one of their date elements, as follows;

- Instead of:
2009-04-09T00:00:00+10:00
- We had:
2009-04-09T00:00:00+10:00:00

This seemingly innocuous formatting error meant that the VEOs could not progress to permanent storage in the Digital Archive.

The Solution

At first glance, the solution to these problems seemed relatively straightforward. A VEO in one sense is just a text file and as such can be edited – this meant that the unwanted colon and zeros could simply be deleted. However, 56,000 VEOs are too many to fix manually and the act of altering each VEO would break its digital signature, which in itself would prevent storage in the Digital Archive.

In light of this, we had to find a way to edit each VEO without an operator having to manually open each file and perform a search and replace operation. We then had to find a way to digitally re-sign them.

“We had to find a way to edit each VEO without an operator having to manually open each file and perform a search and replace operation...”

We used a free, open source text editor known as ‘Sed’. Sed is a (s) tream (ed)itor (hence the name), meaning that it can read a text file line-by-line, finding matches for

VEO - VERS Encapsulated Object

A record which has been encapsulated using XML as outlined in *PROS 99/007 Specification 3*.

Conforms to the VERS metadata scheme as outlined in *PROS 99/007 Specification 2* and contains documents expressed in a long term preservation format.



PROV Digital Archive

The PROV Digital Archive was developed for Victorian Government to manage, preserve, and enable online access to permanent value VEOs.

The access component to the PROV Digital Archive also provides user access to the physical collection at PROV. The PROV Digital Archive also includes a digital repository facility.



for patterns of text within the text stream.

According to the particular commands passed to it by an operator, it can also perform operations on the text. The great advantage of Sed is that you only have to tell it what to do once and from this the program will then work its way through any number of files automatically, making whatever edits it is instructed to make.

To fix the date format error, we instructed Sed to look for the text pattern ':00:00' and replace it with the text ':00'. We then created a series of commands to identify and strip all the digital signature data, leaving only the data enclosed within the VERS Signed Object Element within the VEO. DOS batch files, text files containing a list of commands, were then written to help streamline the process.

“The actual editing and re-signing was completed within a matter of hours - much less work than having to re-create the VEOs from scratch...”

To re-sign the VEOs, we used a java based tool called VEOSigner, which PROV had previously developed in-house. The only problem was that VEOSigner was originally developed to process one file at a time, so we had to write a DOS batch file to automate the process.

Whilst writing and checking the editing commands and batch files took some time, the actual editing and re-signing was completed within a matter of hours - much less work than having to re-create the VEOs from scratch.

Digital Signature

A security mechanism that demonstrates that a particular piece of data (e.g. a record) has not been altered since creation.

i

Sed

Sed is a stream editor. A stream editor is used to perform basic text transformations on an input stream (a file or input from a pipeline). Sed is typically used for extracting part of a file using pattern matching or substituting multiple occurrences of a string within a file.

i

For further information on this process and/or sample Sed and DOS commands please **contact us**.

For further information about Victorian Electronic Records Strategy please go to:
<https://prov.vic.gov.au/recordkeeping-government/vers>

Cover image source: "board-electronics-computer" by blickpixel / CC0 1.0.

Last updated April 2020 (updated hyperlinks)

