Public Record Office Victoria
Advice to Victorian Agencies
July 2003, Version 2.0

Advice 9

Introduction to the Victorian Electronic Records Strategy (VERS) PROS 99/007 (Version 2)

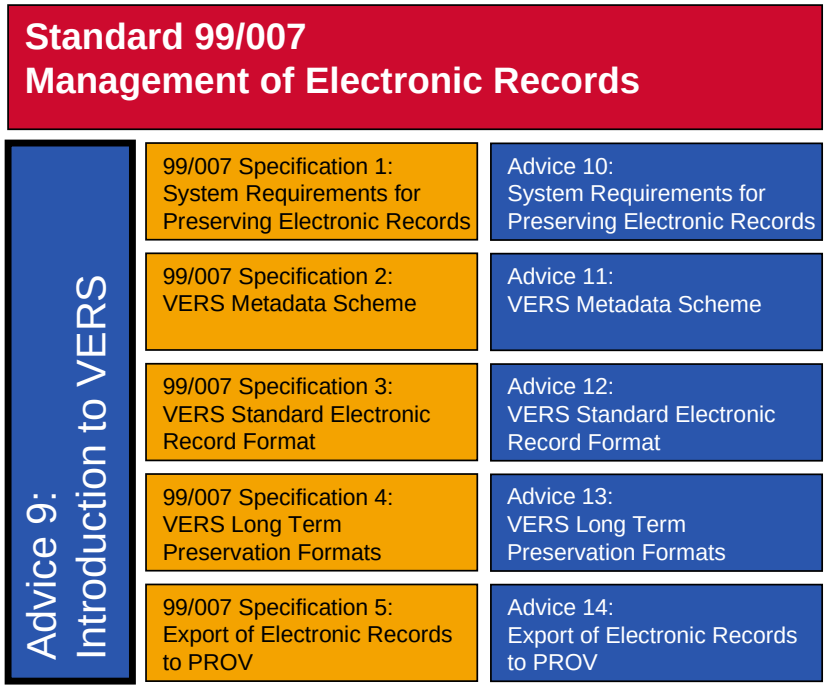
State Government Victoria
Department for Victorian Communities

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

Version	Version Date	Details
2.0	31 Jul 03	Released

The Victorian Electronic Records Strategy (VERS)

This document is an introduction to the PROV Standard 'Management of Electronic Records (PROS 99/007)', also known as the VERS Standard. 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:

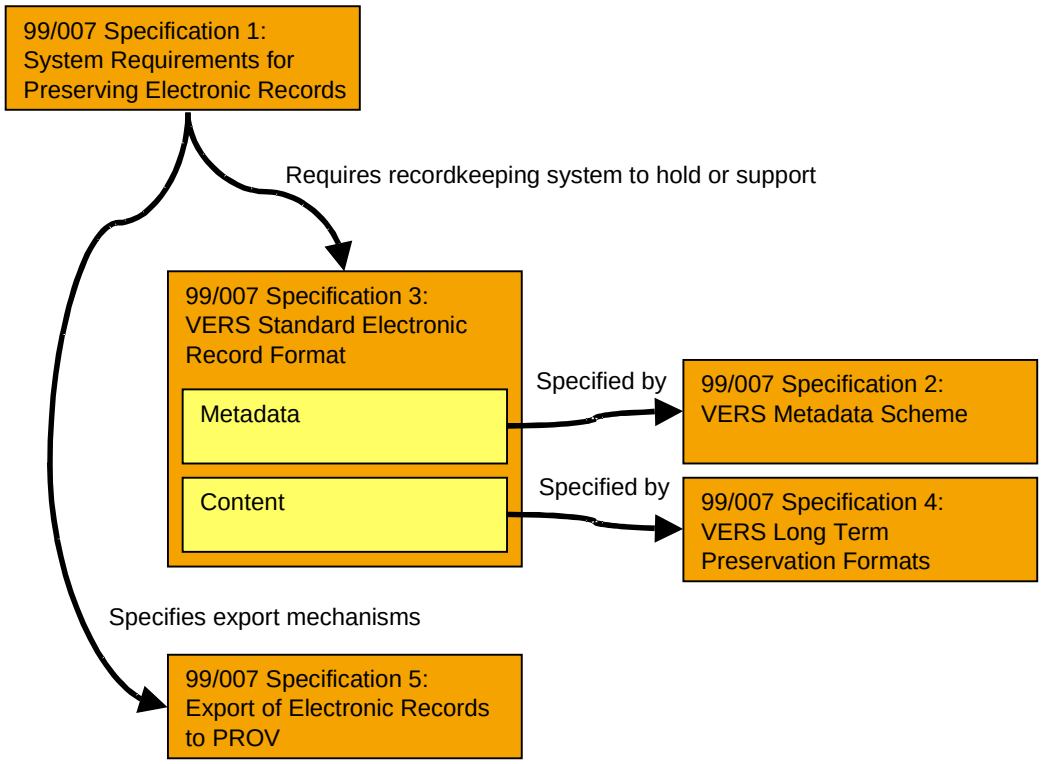
- *Management of Electronic Records*. This document is the Standard itself and is primarily concerned with conformance. The technical requirements of the Standard are contained in five Specifications.
- *Introduction to VERS.* This document provides background information on the goals and the VERS approach 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

1 Introduction 7

2 History of VERS 7
2.1 The 'Keeping Electronic Records Forever' report 7
2.2 The Victorian Electronic Records Strategy report 8
2.3 The VERS Standard 9
2.4 The VERS implementation at the Department of Infrastructure 9
2.5 VERS Centre of Excellence 10

3 Goals of VERS 11

4 Preservation Approach 12
4.1 Program obsolescence 12
4.1.1 The challenge 12
4.1.2 VERS approach 13
4.2 Loss of context, authenticity and integrity 15
4.2.1 The challenge 15
4.2.2 VERS approach 17
4.3 Media refreshing 18
4.3.1 The challenge 18
4.3.2 VERS approach 19
4.4 Loss of records 20
4.4.1 The challenge 20
4.4.2 VERS approach 20
4.5 Loss of recordkeeping system 20
4.5.1 The challenge 20
4.5.2 VERS approach 21

5 VERS Implementation 21
5.1 Requirements on recordkeeping systems 21
5.1.1 Functions 22
5.1.2 Native versus export compliance 23
5.2 Exporting records to PROV 24
5.2.1 Conversion of content to a long-term preservation format 24
5.2.2 Metadata 25
5.2.3 Encapsulation into VERS Encapsulated Objects (VEOs) 25
5.2.4 Export to PROV 26

6 VERS Encapsulated Objects 26
6.1 Record structure 27
6.1.1 Documents 27
6.1.2 Encodings 29
6.2 Record, Document, and Encoding metadata 30

7 References 31

1 Introduction

The Victorian Electronic Records Strategy (VERS) addresses the cost-effective, long-term, preservation of electronic records.

The structure and requirements of VERS are formally specified in the Standard for the Management of Electronic Records (PROS 99/007) and its five technical specifications.

This Advice provides background to the Standard. It covers

- the history of the VERS project
- the preservation theory behind VERS
- how the five specifications support the preservation theory
- a brief introduction to the VERS Encapsulated Object (VEO).

In this document we distinguish between the record and the content of the record. The content is the actual information contained in the record; for example, the report or the image. The record as a whole contains the record content and metadata that contains information about the record, including its context, description, history, and integrity control.

2 History of VERS

2.1 The ‘Keeping Electronic Records Forever’ report

The development of VERS began in 1994 when Public Record Office Victoria (PROV) realised that it was facing a significant challenge in the preservation of records being produced by Victorian Government agencies. The work of the Victorian Government, like most large organisations, was beginning to be carried out electronically. Documents were largely produced electronically using office applications and exchanged using email. Documents were often only printed for proofreading, reading away from the desktop (e.g. in meetings), or for filing as records in conventional paper-based recordkeeping systems. It was realised that many records were not being printed and filed. Instead, records were being kept electronically on shared file systems and in applications, particularly email.

It is interesting to note that at the time the focus of electronic recordkeeping was on the internal processes of government; the external interface with the public was still being handled largely by paper. A major change in the environment since 1996 is the growth in the provision of services to the public by electronic mechanisms, particularly via email and the Web. This has brought an entirely new area of electronic records to the fore.

In 1996 PROV was provided with $240,000 funding from the Microeconomic Reform Fund and retained Ernst and Young Consulting, in conjunction with the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Professor Michael Vitale of The University of Melbourne, to solve the problem of ensuring the retention of, and ongoing access to, electronic records created today and in the future. The terms of reference for the project were:

- to develop, in consultation with experts from other Australian archival institutions, a strategy for the management of the Victorian Government's electronic records, including:
  - the transfer, storage, preservation, access and disposal of these records; and
  - the provision for records to be maintained as evidence.
- to examine possible information technology solutions to support the electronic records management and archive strategy.
- to examine the viability and costs of both a distributed custody/network model and a centralised storage model for the management, storage, and access to electronic records.
- to examine the consequences of any recommended strategy for electronic records for the ongoing management of paper records.

Sections of the archival community, in particular, had been aware of the challenges of preserving electronic records since the early 1990s. Much of the published work was concerned with highlighting the challenges of preserving electronic records or exploring underlying issues of archival theory. Valuable as this work was, little of it was concerned with proposed solutions. Ernst and Young could not identify any tested solutions to the challenges. However, the report drew on existing work, particularly the Pittsburgh project [PITT], and identified encapsulation as a possible solution to the challenge of preserving electronic records.

The result of this project was the report 'Keeping Electronic Records Forever' [PROV1], published in 1996. The report advocated that, instead of taking a system-orientated approach to electronic records, a data-driven approach was more appropriate, as the records would need to outlast any system developed to manage them.

2.2 The Victorian Electronic Records Strategy report

In 1998 the second stage of VERS commenced. The Victorian Government, through the Victorian Microeconomic Reform program, funded a technical trial of a data-driven approach to preserving electronic records. The goal of the project was the construction of a demonstration records management system that captured, encapsulated, and managed electronic records. This involved:

- replicating existing information from PROV and the Department of Infrastructure (which was selected as the trial partner agency for the project)
- determining record capture and access points within the agency's key processes
- adapting technology to provide 'live' experiential testing of the proposed solution
- performing experiments to provide PROV with key sizing and costing parameters for future implementations.

This stage was contracted to CSIRO, who subcontracted Ernst and Young to carry out process modelling within the Department of Infrastructure.

An emphasis of the project was on the capture of electronic records automatically from workflows. This reflected a recommendation of the original report that electronic records should be captured into a recordkeeping system as soon after creation as possible. Although successful in capturing high-quality records with little effort from users, this aspect of VERS was subsequently de-emphasised, as few processes in government were sufficiently well defined and important enough to justify the cost of constructing automated workflow. The emphasis in subsequent projects swung to the adhoc capture of records.

However, a significant benefit of the interaction with the Department of Infrastructure during this project was an emphasis on the joint recordkeeping requirements of both agencies and the archive. VERS was not envisaged as just a strategy merely for an archive, but something that would aid agencies in their use of records. The key was an understanding that in order to ensure the preservation of electronic records, it was necessary to ensure that government agencies were creating records with appropriate content and metadata. This emphasis was given greater weight during a subsequent phase of VERS when a pilot system was implemented in the Department of Infrastructure.

The project successfully demonstrated the encapsulation of electronic records and set the technical basis for VERS. It was during this project that the technical approach used by VERS was developed. This included the development of the VERS Encapsulated Object (VEO) and the initial selection of a long-term preservation format.

The result of this project was the report 'Victorian Electronic Records Strategy' [PROV2], released in 1999. The report concluded that:

- the capture of electronic records into a long-term format is possible and achievable with current technology
- the archiving of electronic records is possible and achievable now.

2.3 The VERS Standard

The work undertaken in producing the Victorian Electronic Records Strategy Report formed the basis of the first version of the VERS Standard. This Standard, 'Management of Electronic Records' [PROV3], was formally launched by PROV in April 2000.

One major difference between the recommendations of the VERS Final Report and the published requirements of the Standard. This was a reworking of the metadata that had to be captured by recordkeeping systems and included in the VERS Encapsulated Object. Towards the end of the technical investigation documented in the Report, the National Archives of Australia had released their metadata standard [NAA]. It was decided to adapt this metadata standard for a significant portion of the VERS metadata. This was primarily a pragmatic decision based on a decision that Australia was too small a market to support competing electronic records metadata standards.

2.4 The VERS implementation at the Department of Infrastructure

The work undertaken as part of the production of the Victorian Electronic Records Strategy report had demonstrated that a data-driven approach to the preservation of electronic records was technically feasible. The work had not demonstrated, however, that the proposed approach could be economically implemented within an agency.

Accordingly, the next stage was the implementation of a pilot system within an agency. This was intended to develop knowledge about the practical issues involved in creating and capturing electronic records within an agency. Such issues included:

- *The integration with existing IT systems within agencies*. This would include general office applications (such as email and office automation) and special purpose custom applications.
- *The cultural issues involved in deploying an electronic recordkeeping system to all staff*. This would particularly include the usability of the registration system and the records access system.

A key goal of this work was to determine how to assist the agency in capturing quality electronic records. A second key goal was that the practical issues of implementing a fully electronic records management system within a government agency would test and validate the VERS Standard.

To test the organisational feasibility of VERS, the Victorian Government committed a further $4.8 million in 2000-2002 to implement an electronic records management system based around VERS within a medium-sized Victorian agency. The agency chosen was the Department of Infrastructure (DoI). The successful tenderer for the work was Solution 6 (now Alphawest), and staff from the Department of Infrastructure, PROV, and CSIRO were involved.

The focus of the VERS@DoI project was the adhoc capture of records whilst the employees of the agency conducted their normal day-to-day business. In implementing such a system, a number of important lessons were learnt about the capture of quality electronic records. These lessons included:

- the importance of tying the recordkeeping system to wider corporate knowledge management goals
- the limited amount of time users are willing to spend learning how to use a new system that is not directly related to the tasks they perform
- the range of ways in which users organise information into records and files.

In addition to these corporate lessons, valuable lessons were learnt about the VERS Standard itself. These lessons included:

- that additional explanatory material was necessary to assist in implementations of the Standard
- areas where the Standard could be extended
- several (minor) areas in which the Standard contained inconsistencies.

The VERS@DoI system went into production use in 2002, and demonstrated that it was possible to implement a fully electronic records management system within an agency.

2.5 VERS Centre of Excellence

As a consequence of the successful implementation at the Department of Infrastructure, the Victorian Government committed $8.2 million over two years to fund a ‘VERS Centre of Excellence’ within PROV. The Centre commenced its work on 1 July 2002.

The Centre has two tasks. The first task is to support Victorian agencies in obtaining and implementing VERS-compliant recordkeeping systems to capture and manage electronic records. Activities undertaken as part of this task include:

- outreach efforts to raise the awareness of senior public sector employees as to the benefits and requirements of electronic recordkeeping
- consulting to government agencies on their electronic records needs and how to make their existing and proposed electronic recordkeeping systems VERS-compliant
- formal training on electronic records principles and practices
- revising the VERS Standard
- extending our understanding of preserving electronic records, particularly capturing additional contextual information available in modern electronic recordkeeping systems

- advising the Victorian Government on the legal, regulatory and policy implications of the electronic environment and providing input into whole-of-government directions in this area.

The second task is to implement a digital repository to manage and preserve electronic records at PROV as part of the permanent archive of the State of Victoria. This archive is expected to be operational in 2005.

A major activity of the Centre is the revision of the original VERS Standard to take into account the experience with implementing and using VERS. Interestingly, the technical component of the Standard has barely changed, although the opportunity has been taken to correct a small number of flaws in the original Standard, and to be more precise in some requirements.

The new version now emphasises cost-effective preservation. In the first Standard, the challenge was to demonstrate that electronic records could be preserved at all. In the revised Standard the ability to preserve of electronic records is assumed; the challenge is to carry it out in such a way that it is economically feasible for both an agency and PROV.

A second emphasis of the revision is to focus and improve the clarity of the Standard. It is hoped that this will assist both vendors and agencies in implementing VERS.

3 Goals of VERS

The goal of the Victorian Electronic Record Strategy is the cost-effective long-term preservation of electronic records.

As the State archive, PROV is primarily concerned with the preservation of records of permanent value to the State of Victoria. These records will eventually come into PROV custody. The focus of the Standard is therefore on:

- the functions that recordkeeping systems in agencies must support in order to preserve electronic records whilst they are being held by the agency
- the physical representation of the records when they are exported from an agency to PROV
- the mechanisms used to reliably export records from an agency to PROV.

However, it is recognised that agencies holding long-term temporary records will face the same preservation issues as those faced by PROV. PROV consequently recommends that the principles underlying VERS (and many of the requirements and techniques used within VERS) should be used by agencies holding long-term temporary records. A ‘long-term’ electronic record is considered to be one that:

- *Will outlive the system that is currently holding the record*. Experience has shown that systems holding electronic records will be replaced every five to ten years. A temporary record with a retention period of 90 years (e.g. medical records) could consequently be expected to be held by between nine and eighteen systems over its life.
- *Will outlive the agency that is currently holding the record.* Some agencies, such as Royal Commissions, may have a life-span measured in months. Others, such as the major courts, have a lifespan of over 100 years.

It should be emphasised that records are not only held in formal recordkeeping systems. Many agencies hold records in business applications (for example, plan management

systems or a case management system). Such applications are not usually considered to be ‘recordkeeping’ systems, but the requirements of VERS apply to such systems just as much as they apply to formal recordkeeping systems.

A major change in this version of the Standard has been the emphasis on cost-effective preservation. Poor choice of preservation techniques will add significantly to the cost burden of the agencies or the archive. A particular challenge is to divide the cost burden for preservation between the agency and the archive in an equitable manner.

As an approach, VERS has always been pragmatic. The team developing VERS has always recognised that electronic records are being created now and that it is far better to deploy an adequate solution to allow the preservation of these records than to lose records by waiting until the perfect preservation approach has been developed. Furthermore, it was recognised that perfection can only be achieved through operational experience of preservation systems in both agencies and in archives. This operational experience can only be gained through deployment of systems. VERS has been extremely fortunate in having extensive opportunities to work with users and recordkeeping staff in agencies.

4 Preservation Approach

There are five main challenges when preserving electronic records over a long period. These are:

- program obsolescence
- loss of record context, authenticity and integrity
- media failure
- reliability
- loss of recordkeeping system.

4.1 Program obsolescence

4.1.1 The challenge

The key challenge to long-term preservation is preserving the ability to render the information contained in a digital file.

This challenge arises because the information in a digital file is simply a sequence of binary data: ‘ones’ and ‘zeros’. Unlike writing, the sequence has no inherent meaning: a sequence could be a report, an image, a musical work, a database, a computer program or anything else. To render the information it is necessary for a program (or application) to interpret the binary data.

Most computer users are already familiar with this problem, as they would have had the experience of being emailed an attachment that they cannot open because they do not have the necessary program installed on their computer. In order to display the attachment it is necessary to:

- work out what the format is
- identify an appropriate program

- either install the program on the computer or get the sender to resend the attachment in a form the receiver can render.

These basic steps are also necessary to render an electronic record. Any solution to the preservation of electronic records encompass allow the identification of the format and the acquisition of an appropriate program to display it. Over time, a major difference with electronic records is that the creator of the record is unlikely to be available, nor will the software that was used to create it. It is consequently not always possible to resave an aged archival record in another format.

The unfortunate problem with obtaining a program to display a particular format is that programs are inherently fragile. They depend for their correct operation on a complicated computer infrastructure. This infrastructure includes the hardware of the computer, the operating system, supporting tools such as compilers or interpreters, software libraries, and even the organisation of computer files in the file system. If any of this infrastructure is changed, a program may cease to function, which, in turn, will make the records rendered by that program inaccessible. Change in infrastructure is inevitable as computer technology develops.

There is also a commercial aspect to the ability to render the information contained in a digital file. It is normally necessary to purchase the programs used to render the information contained in a digital file, and, as the infrastructure changes, to purchase upgrades or new versions of the programs. This is a continual cost to providing access to the digital information. Further, upgrades can only be obtained whilst the vendor continues to support the program. Over 100 years or more, it is reasonable to assume that support for most current programs will cease, either by the vendor going out of business or due to a commercial decision by the vendor to cease supporting the product.

One final issue with program obsolescence is the accuracy of rendering the information. Programs interpret the digital data. If the interpretation changes or is incorrect, the rendering of the information will change. Again, most users will have received an email attachment that is displayed incorrectly because they are using a different program to that used by the sender when creating the attachment. Alternatively, the sender and receiver are running different versions of the same program and the two versions do not render the program in the same way.

4.1.2 VERS approach

The approach taken in VERS to solve the problem of application obsolescence is the conversion of the record content to a long-term preservation format[1]. The long-term preservation format is chosen to minimise (ideally to avoid) the problem of application obsolescence. The value – or otherwise – of this approach depends on the selection of an appropriate preservation format. The long-term preservation formats accepted by PROV are listed in *PROS 99/007 Specification 4: VERS Long-term Preservation Formats*.

Conversion to a well chosen long-term preservation format reduces or avoids the problem of application obsolescence by allowing the rendering program to be re-implemented from scratch, if necessary, or allowing the record to be subsequently converted to a replacement format. We refer to this approach as a ‘data centric’ approach; the focus is on the format of the data. It is to be contrasted with an ‘application centric’ approach, where the focus is on preserving the applications (programs) that access the data.

[1] VERS allows, and encourages, users to also preserve a copy of the record content in the original format in the record. This gives far more flexibility in using the record, while the original format remains accessible.

Ideally, the long-term preservation format allows a rendering program to be re-implemented from scratch in the future. To allow this it is necessary for the data format to be accurately specified.

In choosing appropriate formats, VERS uses the following criteria:

- *Simple format*. The ideal preservation formats are those that are sufficiently simple that it is possible to include a complete specification of the format with each record. Such a description would have to be short: no more than a hundred words or so. Very few formats are this simple, but an example might be a scientific data file which consists of a table of integers.
- *Published formats*. The more common situation occurs where a data format is defined by one or more published specifications. There are many such formats. The simplest example is a plain text file where the format is defined by a specification such as Unicode (ISO 10646) which defines the character glyphs, character numbers, and the encoding of the character number in the data file. More complex examples of published specifications include the standard image formats such as GIF, TIFF and JPEG. Some published specifications are very complex, including page description formats such as Adobe's PDF.

  Some of these formats are formal de jure Standards published by standards bodies (e.g. JPEG, ISO 10646), while others are de facto standards (e.g. GIF, TIFF, and PDF), which may be proprietary formats. The important feature of all of these formats is that the specification is published, is available, and will be continue to be available for the indefinite future. A conservative archive should, of course, obtain reference copies of the specifications for the data formats it accepts.

  Formal de jure standards are preferred as long-term preservation formats, however, because it is more likely that vendors will implement them accurately. The problem with proprietary formats, particularly those where only one or two implementations exist, is that the vendor that owns the format may 'cheat' and either not implement the format accurately, or add additional undocumented features.

There are often several suitable published formats which may be chosen as a long-term preservation format. In this case, consideration should be given to what characteristics of the record it is important to preserve over a long period of time. For example, PROV has judged that a key characteristic of record it was necessary to preserve was the appearance of the record as the original creator saw it. This led to the selection of PDF as a long-term preservation format over an XML format, as PDF can ensure a far more accurate representation.

Where there is no suitable published format, VERS recommends choosing a widely used industry standard format. Perhaps the best example of such a format is Microsoft Word in the word processing arena.

When adopting an industry standard format, a different strategy for long-term preservation must be used. The strategy is to ultimately convert the records from the industry standard format. The organisation holding the records must monitor the availability of software that can render or convert the format, and when the format is becoming obsolete undertake the conversion.

The advantage of adopting an industry standard format is that an archive can harness economics to its benefit. A very widely used industry standard format is unlikely to become obsolete rapidly. Any new program that competes with the industry leader has to convert the data formats used by the industry leader; otherwise the new competitor will be unable to enter the market. Finally, there are likely to be several options for conversion, allowing an archive to minimise cost and maximise the accuracy of the conversion.

Published formats are preferred to unpublished industry standard formats, for two reasons:

- There may be only a short window of opportunity for conversion before an obsolete format becomes unreadable. An archive must monitor the obsolescence of the formats and fund conversion before this window closes.
- The conversion is dependent on externally sourced products and may not be sufficiently accurate for archival purposes.

Conversion to a long-term preservation format is a conversion process, similar to digitising or microfilming paper records, and an agency or archive must ensure accuracy of conversion. For example, there are many methods of converting to PDF, but some of them can produce inaccurate representations of the record. The mechanisms used in microfilming or digitising (e.g. statistical sampling of the conversion process) can be used in ensuring accuracy of digital conversion.

The timing of the conversion has a bearing on the accuracy of conversion. Where the record is converted sometime after it is created, the conversion accuracy may be limited. This may occur, for example, if the conversion program is upgraded and this changes the results of the conversion. In selecting a conversion process, it is worth considering whether the process is used for day-to-day business activities, as this vastly improves the conversion accuracy. For example, PROV has found that the most accurate conversion tool for PDF is Adobe's Distiller. A major reason for this is that the basis for this tool is the use of the standard printing functions in the application producing the PDF. Since printing is a business-critical function, the distillation has a very sound basis for conversion.

4.2 Loss of context, authenticity and integrity

4.2.1 The challenge

There is an extensive treatment of the concepts of context, authenticity, reliability and integrity of electronic records in the archival literature.

The *context* of a record equates to how the record relates to other records held by an organisation. Context is critical to the use of a record. Frequently, an answer to a question will not be given by one record. Instead, a user is interested in understanding a story which is documented in a collection of related records. The context of a record allows the discovery of these related records.

An *authentic* record is one that is capable of being proved to be what it purports to be (i.e. the content is what it appears to be, it was created by the person who appears to have created it, and it was created at the time it appears to have been created).

A *reliable* record is one that contains a full and reliable representation of the facts which the record documents. Note that a record can be authentic, but not reliable. A record is not reliable, for example, if the author of the record left out material facts, misrepresented the position, or simply lied. Such a record would still be authentic as the content is as the author intended and it was created by the apparent author at the apparent time. Authenticity is concerned with the truth of the record as an object; reliability is concerned with the truth of the contents of the record.

Integrity refers to the record being complete and without unauthorised alterations. Note that records can be altered and retain their integrity, provided the alterations are allowed by the policy of the organisation, are authorised, and are documented.

These properties (context, authenticity, reliability, and integrity) are independent of whether the record is paper or electronic. In both paper and electronic records these properties are not contained in the content of the record. Instead, they are partially represented by information associated with the record content (this information is normally known as 'metadata' when dealing with electronic records). Authenticity, reliability, and integrity are also partially dependent on the processes used to capture and manage the records.

The challenge in preserving electronic records is ensuring that the systems that manage the electronic records hold sufficient metadata and implement suitable processes to ensure the long-term retention of context, authenticity, reliability, and integrity.

In a traditional paper-based recordkeeping system these properties are largely demonstrated by the procedures involved in the creation, storage, and handling of the record. For example, reliability is shown by the fact that the record was created for future reference as part of a standard business procedure. Authenticity and integrity is shown by the procedures involved in managing and controlling access to records. Ultimately, these procedures are backed up by conventional forensic tests such as tests on signatures, the age of the paper, type of typewriter, and ink.

This reliance on procedures can be transferred to many electronic records, particularly those managed by application specific systems. Consider a financial system, for example. The records would be considered reliable because they are automatically generated by the system as a side-effect of carrying out financial tasks. They are authentic because the actions can only be carried out via the financial system and the system keeps logs of who carried out the task, when it was carried out, and how the tasks are related. Finally, the logs record any changes to the records, and hence the records have integrity.

However, many electronic records are not managed in such a formal way. This particularly applies to those records held in generic software applications (e.g. email systems) or in the general file system. Fundamentally, the problem is that these systems are not designed to ensure authentic records or to ensure their integrity once created. These records can be the most important held by an agency; for example, they may document the development of government policy.

One method of ensuring authenticity and integrity of these records is to install an application that is designed to manage records and to ensure their authenticity and integrity (a recordkeeping system). Once records are registered with the recordkeeping system, the system can ensure that the record retains integrity. Essentially, the recordkeeping system acts as a vault, mediating and recording access to the records. Just like the financial system, the recordkeeping system only allows certain operations on the registered records, only allows authorised users to perform those operations, and keeps audit trails of all operations.

However, there are several issues with using a recordkeeping system to ensure the reliability and integrity of records.

The effectiveness of a recordkeeping system depends on users placing their records under the control of the system. At some point, for example, users must move their emails from their mailbox to the recordkeeping system. This is to be contrasted to a financial system, for example, where the system is used to carry out the tasks associated with managing money, the records being automatically generated as a side-effect. With a recordkeeping system, the tasks are carried out in other applications and users have to consciously decide to place the records under the control of the system.

Care needs to be taken that users with special access cannot subvert systems holding records. Typical special access users are records administrators or (computer) system administrators. However, it should be noted that such users can equally subvert traditional

paper-based records systems, so this issue is no different in the electronic environment. The question is whether advantage should be taken of technology to close this hole.

Management by a recordkeeping system should be viewed as a medium-term solution. Any computer system has a relatively short life – say five to ten years – and there must be a plan to extract records from a system and to migrate them to a replacement system (or to manage them by some other mechanism if there is no replacement system). This migration is likely to be complex, as it is necessary to preserve sufficient information to show that the record was properly managed to ensure authenticity and integrity when under control of the original system. A particular concern about migration is that this may have to occur under extreme time or budgetary constraints. These constraints typically occur if an agency (or section) is closed and the records are no longer considered of operational interest. An example would be a Royal Commission. Funding for migration is likely to be minimal in these circumstances, and the time available for migration very short.

4.2.2 VERS approach

VERS defines a standard set of metadata that holds the information necessary to show the context, authenticity, reliability, and integrity of a record. The metadata is based upon that defined by the National Archives of Australia [NAA]. The VERS Standard requires that this metadata be encapsulated with the record content in a single object (the VERS Encapsulated Object, or VEO) upon export to PROV. The full VERS metadata is defined in *PROS 99/007 Specification 2: VERS Metadata Scheme*. *PROS 99/007 Specification 3: VERS Standard Electronic Record Format* defines the standard format of the VERS Encapsulated Object and the standard representation of the metadata.

A significant benefit of specifying a standard metadata scheme is in enforcing data normalisation. An archive will receive records from many agencies. This potentially leads to very serious problems of consistency of metadata. It would be almost impossible to provide a unified view of the collection if, for example, each agency used a different metadata element to contain the title. One benefit of defining standard metadata is that the agency performs the normalisation of the metadata. The agency understands the record and the source recordkeeping system.

A second benefit of this approach is that the metadata is encapsulated with the record content in a single object (the VERS Encapsulated Object, or VEO). The importance of this is that it is far less likely for the metadata to become separated from the record content. This is to be contrasted to the situation where the metadata is held in a database separate from the content. In this situation it is easily possible to lose the metadata, or to lose the linkage between the metadata and the content. If either of these situations occur, the record context, authenticity, reliability, and integrity are lost.

VERS uses digital signatures to show that a record has not been altered. A digital signature is the result of applying a mathematical function to the record and is a secret known only to the signer. A related mathematical function can be used to verify the digital signature. The VERS Standard contains metadata elements that contain the necessary information required to validate digital signatures. The way digital signatures are applied to VEOs is defined in *PROS 99/007 Specification 3: VERS Standard Electronic Record Format.*

Many archives do not specify that it is necessary to digitally sign records. Instead, integrity is shown by custody in an archival system. This has been the traditional approach to showing authenticity and integrity of paper records held by an archive. The reason PROV feels that this is inappropriate for electronic records is that custody was always backed up by forensic tests with paper records. Such tests are in their infancy with electronic records. Further, a digital archive is a far less benign environment than a paper repository and records can easily be altered by software bugs and hardware failures. Such failures can systematically

affect large parts of the collection. It was felt that a verification mechanism independent of the digital archive was desirable.

4.3 Media refreshing

4.3.1 The challenge

When the challenge of preserving digital objects comes up, the first issue that most people identify is the lifespan – or lack thereof – of the media on which the digital object is stored. There are really two related issues here:

- *Media deterioration*. The physical media on which the electronic records are stored will deteriorate over time and eventually become unreadable.
- *Media obsolescence*. Even if the media remains readable, it will eventually become impossible to obtain suitable readers to read the media.

The physical media on which the electronic record is stored will deteriorate over time and eventually become unreadable. Part of the deterioration is due to wear and damage as the media is used, and part of the deterioration is due to ageing of the media. Wear and damage occurs to all media. For example, tape wears as it passes around capstans and read/write heads. Even CDs, which do not suffer wear in the same fashion, suffer damage that causes errors due to physical handling of the CD. There have been many examples of slow chemical changes in media causing problems with longevity. Examples include:

- paper-based records on paper with a high level of acid, which would eventually cause the paper to disintegrate
- magnetic tape in which the binder that holds the magnetic particles to the tape decompose and cause the layers of tape to stick together in the reel
- CDs in which poor manufacturing processes allow the aluminium reflective layer to oxidise.

It is possible to reduce the rate at which media deterioration occurs. Wear and damage can be reduced by careful handling and adjusted readers. Chemical deterioration can be reduced by holding the media in a controlled environment. However, it is not possible to prevent media deterioration. Sooner or later, every piece of media will become unreadable.

Short as the lifespan of a piece of media is, the lifespan of the hardware necessary to read the media may be even shorter. A CD may physically last one hundred years, but it will almost certainly not be possible to obtain a CD reader at that time that will read that CD.

Being physical devices, the readers are also subject to mechanical wear and chemical deterioration. Consider a CD reader, for example. The servo mechanism which controls the position of the laser suffers wear and will eventually be incapable of controlling the position sufficiently precisely to read the CD. The laser that reads the CD deteriorates over time and loses power and will eventually be incapable of reading the CD. Deterioration of the readers will be a more significant problem with modern technology than older technology because of the miniaturisation and complexity of modern technology. For example, it is quite feasible to machine a part for a half-inch reel-to-reel tape drive. It will not be economically feasible to grow and fabricate a semiconductor laser of the precise frequency required to read CDs.

The wear and deterioration of media readers means that they will eventually need to be replaced. Unfortunately, hardware has an economic lifespan, and once an item is no longer manufactured the cost of replacement becomes prohibitive. The lack of spare parts usually results in the cost of repair also being prohibitive. Even if it is possible to obtain, or maintain, the hardware, it is usually difficult and expensive to physically connect it to modern

computers and impossible to obtain the necessary software to drive it. The economic life of a media technology varies significantly. Widely-used media built by many vendors (such as 3.25 inch floppies) have a long lifespan. On the other hand, expensive, proprietary, media technology can have effective lifespans of only a few years. Modern technology is likely to have a shorter economic lifespan than older technology due its greater complexity and miniaturisation.

4.3.2 VERS approach

The only currently practical solution to the limited lifespan of media and the hardware that reads them is periodic replacement of the media. This involves copying the digital information from one piece of media to another piece of media. This practice is referred to as ‘refreshing’. The copy may be to the same type of technology (e.g. the replacement of one hard disk driver with another), or it may involve a transfer to a new technology. Both approaches protect against mechanical or chemical deterioration of the media, while the second also protects against technical obsolescence of the media.

The good news is that media refreshing is a completely solved problem. All IT departments routinely refresh media, and tools and systems to manage and perform refreshing are commercially available to support sites with large amounts of data.

The cost and risk of refreshing is largely dependent on the degree of automation of the process. PROV strongly recommends highly automated systems for managing media for this reason.

The most expensive option is where the media is stored offline on shelves. Refreshing consequently involves manual handling of the media; fetching the original media from storage, loading it into the system, and storing the new media back onto the shelves. It is necessary to take great care in the external labelling and handling of the replacement media, as it is easy to misfile or mislabel media and hence lose records.

Fortunately, media management systems have largely eliminated the need for manual refreshing. It is possible to purchase ‘tape libraries’ or ‘silos’ that replace the manual transfer of media from storage to the readers by robots. The great benefit of these systems is the reduced cost of loading and storing media and the reduced risk of misplacing media.

The first risk of refreshing is that the copy will not be exact; that somewhere in the refreshing process the bit stream will be corrupted. Fortunately, it is easy to perform a bitwise comparison of the data from the original piece of media and the copied data to determine the accuracy of the refresh. This, however, will approximately double the time necessary to perform the refresh, and hence halve throughput.

The greater risk of refreshing is the risk of overlooking a piece of media and hence not refreshing it. This raises the risk that the piece of media will be overlooked for so long that it is no longer possible to read the media because of deterioration of the media, or lack of a suitable reader. This risk is minimised by avoiding manual handling, with the attendant risks of misplacing or mislaying media. Media management systems automatically track the location of media. Some types of tapes record the number of times the tape has been used in a chip in the tape cartridge itself.

4.4 Loss of records

4.4.1 The challenge

Records may be lost by system failure. Such failures can include:

- *Corruption due to failure to accurately copy records from one place to another*. This includes errors in copying from one piece of media to another, or from disk into memory. A particular challenge with preventing this type of failure is identifying all those locations in the computer system where records are copied.
- *Corruption due to failure of indexing*. This may result in the records still physically existing but the recordkeeping system 'forgetting' that the record exists.
- *Hardware failure*. Records ultimately have a physical representation, either on media (e.g. disk or tape), or in memory. Hardware failures such as disk crashes can cause the loss of the record.
- *Disaster*. Records will be lost if the computer holding them is damaged by a disaster such as a fire or flood.

4.4.2 VERS approach

The VERS Standard requires the recordkeeping system *as a whole* to be reliable. In this case the system is not just considered to be the actual recordkeeping application, but includes:

- the hardware on which the application runs
- the system software, such as the operating system and storage management systems
- the processes and procedures that surround the system such as back-up regimes and disaster recovery.

Some failures are due to poor software engineering and can be protected against by ensuring the use of quality software products that have been analysed to identify possible points of failure and have been engineered to guard against them.

Other failures cannot be prevented. It is impossible to prevent hardware failure or a disaster, for example. Such failures must be protected by processes and procedures instituted by the agency and designed to allow recovery of the records. These processes and procedures must be regularly tested to ensure that they work. In many contexts, it would make sense for records disaster planning to be incorporated into the wider organisational disaster-recovery planning.

4.5 Loss of recordkeeping system

4.5.1 The challenge

It does not appear to be widely appreciated that a recordkeeping system itself is another program. It holds records in proprietary data structures; in particular the metadata is often held in database tables separately from the record content. Loss of the recordkeeping system can consequently cause loss of the ability to display the records. The recordkeeping system can be lost for a number of reasons:

- *Abolition of the agency that runs the recordkeeping system*. This may occur with little warning, particularly for short-lived agencies such as Royal Commissions. Usually the staff at agencies about to be closed have other priorities than ensuring the preservation of records.

- *Unplanned termination of service*. This usually occurs because the hardware on which the recordkeeping system runs fails, and it is not considered economic to repair the hardware or relocate the recordkeeping system to another machine.
- *Planned termination of service*. Sooner or later every program will be decommissioned, normally because it is no longer economic to run the application (e.g. due to license fees or because it is not worth relocating the application to a new machine). There should be plenty of warning that this is to occur, but this does not always happen.

For these reasons it is essential to plan for the export of records and to ensure that the records can be extracted from the system with little notice.

The export function must be built into the recordkeeping system from the time of commissioning. Extracting records from a recordkeeping system at short notice before the recordkeeping system is decommissioned is likely to be extremely expensive. Extracting records after the recordkeeping system has been decommissioned will be even more expensive, and may not be practical at all.

4.5.2 VERS approach

The VERS Standard requires the export function to be built into the recordkeeping system to achieve compliance. For permanent records, the VERS Standard requires the system to be capable of exporting records in the VERS Encapsulated Object (VEO) format.

Use of the export function requires the recordkeeping system to be operational. One way of implementing VERS, however, allows the recovery of records even after the recordkeeping system has ceased to function. If the VERS Encapsulated Object (VEO) is generated when the record is registered into the system, this means that it is possible to recover the record independently of the operation of the recordkeeping system. This is because the VEO contains the record content and the necessary metadata (e.g. the context and history) in one object. The VEO can be recovered from the file system without the recordkeeping system being operational.

5 VERS Implementation

The previous section discussed the theory which underlies the VERS approach. This section covers the practical techniques that are used to translate this theory into practice.

Essentially, the VERS requirements have two emphases:

- requirements on a recordkeeping system to ensure that records are properly managed whilst they are being held by the recordkeeping system
- requirements on export to PROV. These requirements cover the specific format required for export and the export processes required.

5.1 Requirements on recordkeeping systems

The VERS requirements on recordkeeping systems are to ensure that the records are properly managed within an agency to ensure that they are complete, documented, and retain integrity.

When considering these requirements, it is important to remember that the life of most records will far exceed the life of any individual recordkeeping system that holds them. This

means that most records will be transferred several times from one recordkeeping system to its successor. Many of the requirements apply to the complete life of the record. For example, it is necessary to document the history of the record from initial registration. This implies that the history must be transferred with the record from one recordkeeping system to its successor.

5.1.1 Functions

The requirements on a recordkeeping system that is required to preserve electronic records for a significant period are listed in *PROS 99/007 Specification 1: System Requirements for Preserving Electronic Records*. They can be summarised as follows:

- *The ability to export records*. As discussed in section 4.4, it is essential that a recordkeeping system be capable of exporting the records (both content and metadata). When exporting to PROV, the recordkeeping system must be capable of exporting the records in the VERS format.
- *Ensuring the integrity of the record*. It must be possible to show that the record has not been modified in an unauthorised fashion (i.e. has retained integrity) since creation. Authorised modifications must be documented in the history associated with each record. This is to prevent the loss of integrity discussed in section 4.2.
- *Documenting the history of the record*. It must be possible to document the history of the record from the time of its creation. This includes the registration of the record, reclassification or alteration to the context of the record, any preservation actions (e.g. format conversions), transfers between recordkeeping systems, and export to PROV. This is to prevent the loss of context, authenticity, and integrity discussed in section 4.2.
- *Documenting the creation of the record* (i.e. being able to show that a record is authentic). This includes documenting who registered the record, when it was registered, and the context of the registration. This is to prevent the loss of authenticity discussed in section 4.2.
- *Metadata capture*. The recordkeeping system must capture and store sufficient information to document the context of the record. This is to prevent the loss of context discussed in section 4.2.
- *Conversion to long-term preservation format*. At some point the record content must be converted to a long-term preservation format to ensure that access to the record is independent of the application that created it. This issue is discussed extensively in section 4.1. Conversion may occur at any time, but two common points are when the record is registered, or when the record is exported. Late conversion increases the risk of record loss, as it increases the chance that the necessary application will not be available or will produce an inaccurate conversion.
- *Reliability*. The recordkeeping system must not lose records entrusted to its care, as discussed in section 4.4. Records may be lost by software failures (e.g. inaccurate copying), or by catastrophes (e.g. the system being destroyed). Reliability is partially addressed by the quality of the engineering within the recordkeeping program; including that defensive coding practices have been applied (these check that functions have worked correctly). However, quality software will not, by itself, be sufficient to ensure that records are never lost. All software must be assumed to contain bugs. In addition, software cannot guard against hardware failures or natural disasters. To prevent the loss of records, the agency responsible for the recordkeeping system must institute and test a proper backup and disaster-recovery regime.
- *Refreshing*. The recordkeeping system must be capable of accurately refreshing the media on which the records are held. This is to cover the mechanical and chemical

deterioration of both the media and the hardware that reads the media, as discussed in section 4.3.

5.1.2 Native versus export compliance

The VERS Standard does not specify the mechanisms by which recordkeeping systems should conform to the requirements it imposes. Different products and systems will satisfy the requirements in different ways, and this is seen as both appropriate and necessary in a market with diverse needs.

However, it is envisaged that there will be two main implementation models. These models are referred to as 'native' compliance and 'export' compliance. Both models require that records be exported to PROV as VERS Encapsulated Objects (VEOs). The models differ in when the VEOs are generated.

With export compliance, the VEOs are only generated when the records are to be exported to PROV. Before the records are exported they are held in the internal format of the recordkeeping system. When using this model, the component that creates the VEOs can be viewed as an additional module that converts the records from the internal format to the VEO format required by PROV. The advantage of this implementation approach is that it requires minimal changes to existing recordkeeping systems.

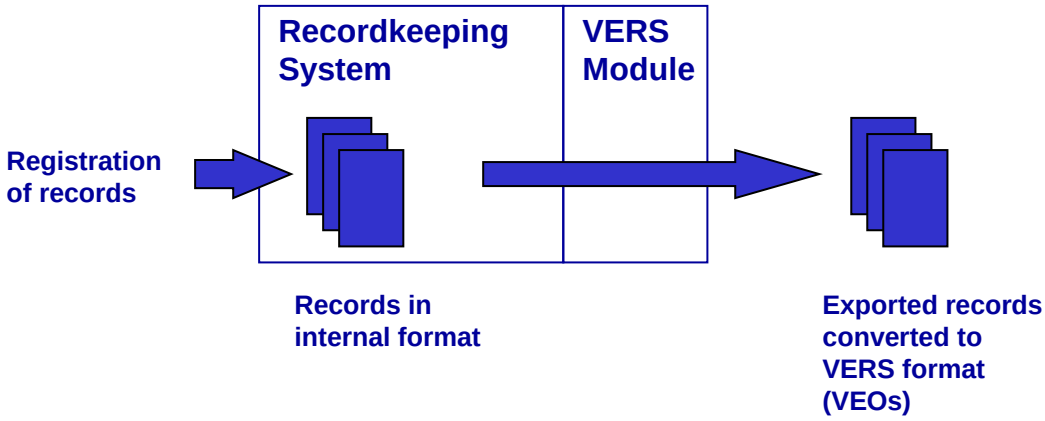


Figure 1. Export implementation model. In this model, records are held within the recordkeeping system and only converted to VEOs upon export.

With native compliance, however, the VEOs are created when the record is initially registered into the recordkeeping system. The recordkeeping system then holds the records as VEOs until they are exported to PROV.

The advantage of native compliance (see Figure 2) is that the record is converted to a long-term format, the necessary metadata is collected, and the record is signed using a digital signature at registration. This means that the record is not dependent on the continued functioning of the recordkeeping system. As discussed in section 4.4, even if the recordkeeping system should catastrophically fail, the VEOs will be sitting on the filesystem and can easily be extracted and transferred to PROV.

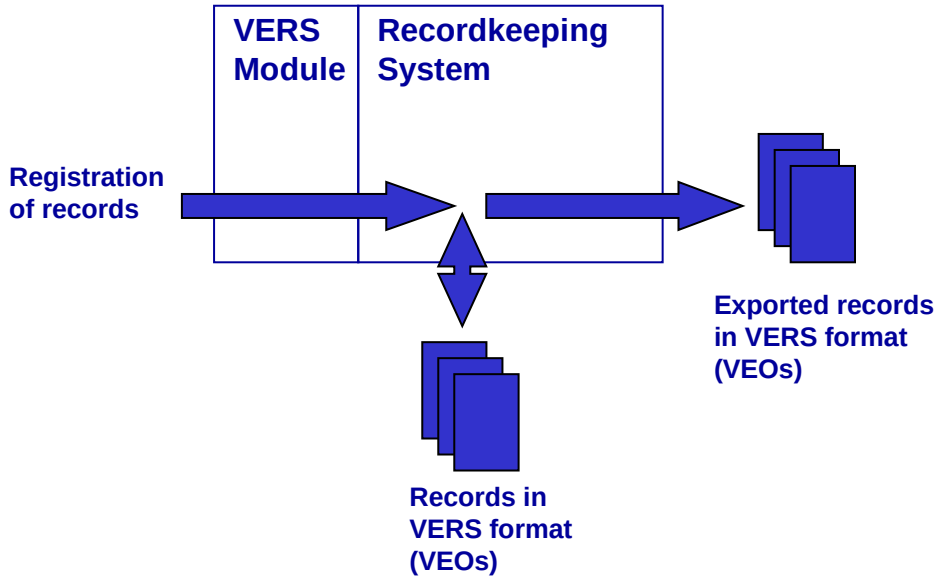


Figure 2. Native implementation model. In this model, records are converted to VEOs when first registered. One advantage of this model is that the VEOs can be directly extracted from the source system if necessary.

5.2 Exporting records to PROV

The VERS Standard imposes rigorous requirements on the export of records to PROV. These requirements deal with:

- the physical representation of the records. This physical representation is known as the VERS Encapsulated Format (VEO)
- the mechanism by which records are exported to PROV.

There are two reasons for these export requirements:

- *Long-term preservation*. The requirements have been primarily formulated to support the long-term preservation of records.
- *Cost-effective preservation*. A secondary reason for the requirements is to reduce the cost of export and preservation to both an agency and PROV.

5.2.1 Conversion of content to a long-term preservation format

At some point in the life of the record at or prior to export, the record content must be converted to one of the long-term preservation formats specified in *PROS 99/007 Specification 4: VERS Long Term Preservation Formats*.

These formats are selected to address the challenge of application obsolescence described in section 4.1. The conversion also has the secondary pragmatic benefit of reducing the cost of providing ongoing access to the records, as each additional format supported by an archival agency requires access software to be purchased or written.

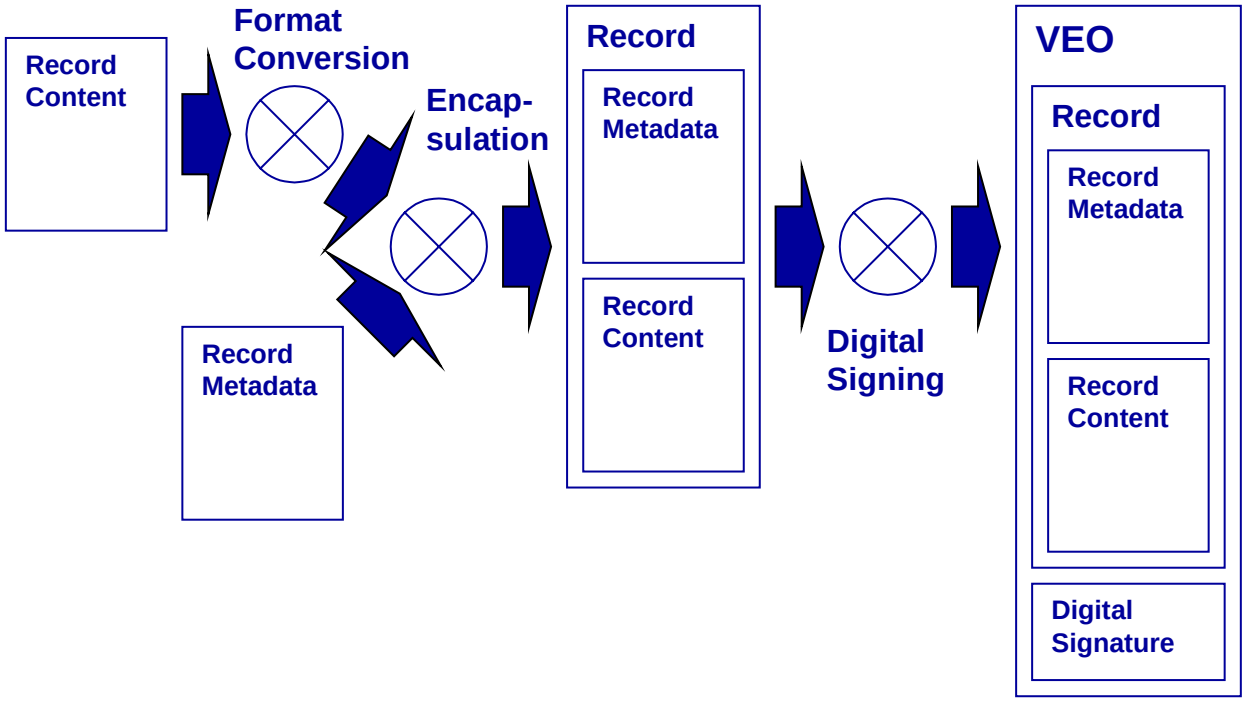


Figure 3. Nominal process of creating a VERS Encapsulated Object (VEO). The record content is first converted to the suitable long-term preservation format. The result of the conversion is combined with the record metadata to form the record. The result is digitally signed to form the VEO.

5.2.2 Metadata

The VERS Encapsulated Object (VEO) contains a very extensive collection of metadata that documents the record's context, history, and technical details. The metadata is specified in *PROS 99/007 Specification 2: VERS Metadata Scheme*. Every VEO must contain the mandatory metadata listed in this specification (and conditional metadata if the condition applies). This metadata prevents the loss of context, authenticity, and integrity described in section 4.2.

Defining a specific set of metadata has two benefits. The first is that the mandatory metadata ensures that each record contains at least a minimal description. The second is that all records transferred to the archive have a coherent set of metadata. A title, for example, is always a title. Consistent use of metadata is essential to allow an archive to provide a unified view of records from hundreds of (different) agencies.

5.2.3 Encapsulation into VERS Encapsulated Objects (VEOs)

The record content and metadata is encapsulated into a single object referred to as a VERS Encapsulated Object, or VEO. The purpose of the VEO is to have a representation of the record that is independent of any individual recordkeeping system.

The use of a VEO has two benefits:

- *Reduced cost for agencies and archives*. As all records are transferred in a common format, the cost of performing the export is minimised. It is hoped that support for VERS will be available in common recordkeeping applications and agencies will be able to produce VEOs and export to PROV without additional development work. Within the archive, it will not be necessary to develop software and processes to deal

individually with each transfer. Once in the archive, each record will have exactly the same structure, which drastically simplifies the management and access systems.

- *Program independence*. A VEO contains all the information necessary to preserve an electronic record. It is consequently independent of the recordkeeping system used to manage the records at any given time. This assists in guarding against the loss of the recordkeeping system described in section 4.5.

The VEO format is defined in *PROS 99/007 Specification 3: VERS Standard Electronic Record Format*.

5.2.4 Export to PROV

The final requirements of VERS detail the mechanisms and processes used to physically export the records to PROV. These requirements are defined in *PROS 99/007 Specification 5: Export of Electronic Records to PROV*. The areas covered by this Specification include:

- *Physical transfer media*. This includes the types of media that will be accepted and any restrictions on individual media types.
- *Archiving software*. This defines how records are to be written to tape media.
- *Acceptance of records*. This describes the protocol by which formal custody of records is accepted by PROV.

A standard export mechanism is essential to minimising cost for both agencies and PROV.

6 VERS Encapsulated Objects

The core of the VERS Standard is the VERS Encapsulated Object, or VEO. This section introduces the overall design of the VEO and gives some background on its technical and archival features. More information on the VEO can be found in *PROS 99/007 Specification 3: VERS Standard Electronic Record Format,* and Advice 12 on that Specification.

One way of thinking of VEOs is by considering them as a message sent from the computer that created the VEO to a second computer in the future (which may not even have been built yet).

VEOs are intended to be self-documenting; that is, in 100 years time a technical user can examine the contents of a VEO and extract sufficient information from it to begin the process of extracting the content. To this end, the VEO has the following features:

- *Single object*. A record is contained within a single computer file. Only that file needs to be examined to understand what the record contains, its relation with other records, and its history. The use of a single file makes it easy to copy the record, and difficult to lose part of the record.
- *Textual content*. The contents of a VEO are plain ASCII text. As such, the contents can be displayed by the simplest computer programs. Examples of these programs are ‘type’ in MS-DOS, ‘Notepad’ and ‘WordPad’ in Windows, and ‘cat’, ‘vi’ and ‘more’ in Unix. We expect such plain text to be readable indefinitely.
- *Textual markup*. Each piece of information (metadata) in a VEO is labelled with a descriptive tag intended to indicate the purpose or function of the information. The tags were chosen to be easy to understand; abbreviations, for example, were avoided.
- *Embedded documentation*. Each VEO contains a number of short pieces of text that describe the technical features of the VEO. These pieces of text provide a summary of

the information necessary to implement a viewer for the VEO and, in particular, reference any external specifications necessary. One of the more complex embedded documentations is the description of the digital signature:

```
The contents of this VEO are signed using the SHA-1 hash algorithm and the DSA
digital signature algorithm. SHA-1 is defined in Secure Hash Standard,
FIPS PUB 180-1, National Institute of Standards and Technology, US Department
of Commerce, 17 April 1995
(http://csrc.nist.gov/publications/fips/fips180-1/fip180-1.pdf).
The DSA algorithm is defined in Digital Signature Standard (DSS), FIPS PUB
186-2, National Institute of Standards and Technology, US Department of
Commerce, 27 January 2000
(http://csrc.nist.gov/publications/fips/fips186-2/fip186-2-change1.pdf).
Details of the public keys are encoded as X.509 certificates in the
vers:CertificateBlock elements. X.509 certificates are defined in "Information
technology - Open Systems Interconnection - The Directory: Public-key and
attribute certificate frameworks," ITU-T Recommendation X.509 (2000).
The signature and certificates are encoded using Base64. Base64 is defined in
Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet
Message Bodies, Section 6.8, Base64 Content-Transfer-Encoding, IETF RFC 2045,
N. Freed &amp;amp; N. Borenstein, November 1996
(http://www.ietf.org/rfc/rfc2045.txt?number=2045).
The signature covers the contents of the vers:SignedObject element starting
with the 'less than' symbol of the vers:SignedObject start tag, up to and
including the 'greater than' symbol of the vers:SignedObject end tag. Before
verifying the signature all whitespace (Unicode characters U+0009, U+000A,
U+000D, and U+0020) must be removed from the text.
```

6.1 Record structure

The VERS Standard does not assume that a record consists of a single computer file with one representation. VERS records may contain multiple documents, each of which may contain multiple encodings.

6.1.1 Documents

In many cases a record contains multiple independent documents. For example, the VERS Standard consists of six core documents: the Standard itself, and five supporting Specifications. If the VERS Standard was encapsulated in a VEO, it is possible to include all six documents (as separate PDF files) in a single VEO.

The use of documents allows users of the VERS Standard great flexibility in determining what information is contained within a record. For example; a correspondence record could contain a copy of the original incoming letter as well as the outgoing response.

Version 2 of the VERS Standard allows documents to be structured where documents can be organised into a hierarchy. This allows the creator of the document to control how the information in a record is organised and presented to future users.

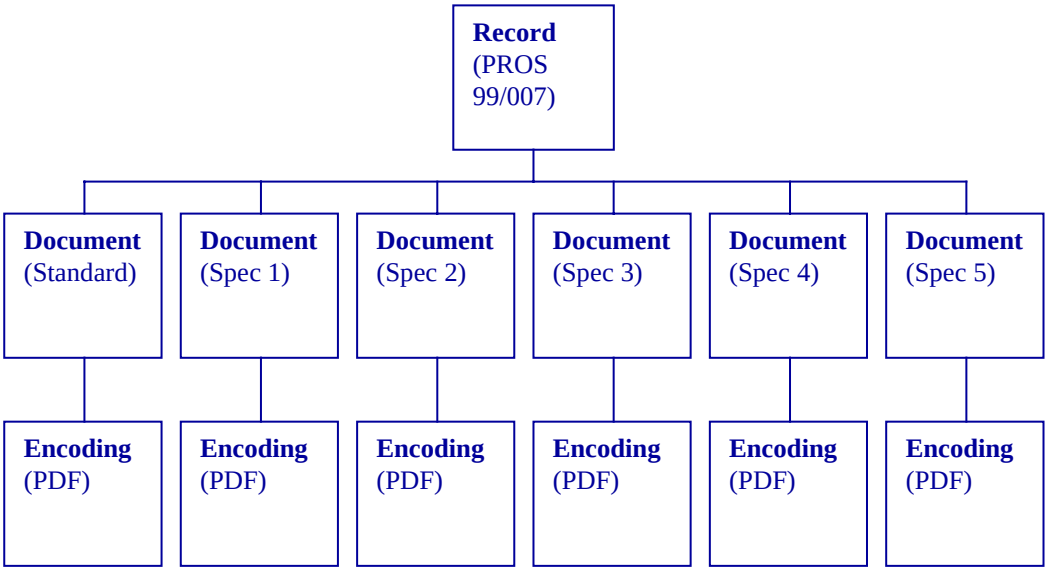


Figure 4. A VERS record of PROS 99/007, which contains six documents: the Standard itself and the five supporting Specifications. Each document is represented by one encoding.

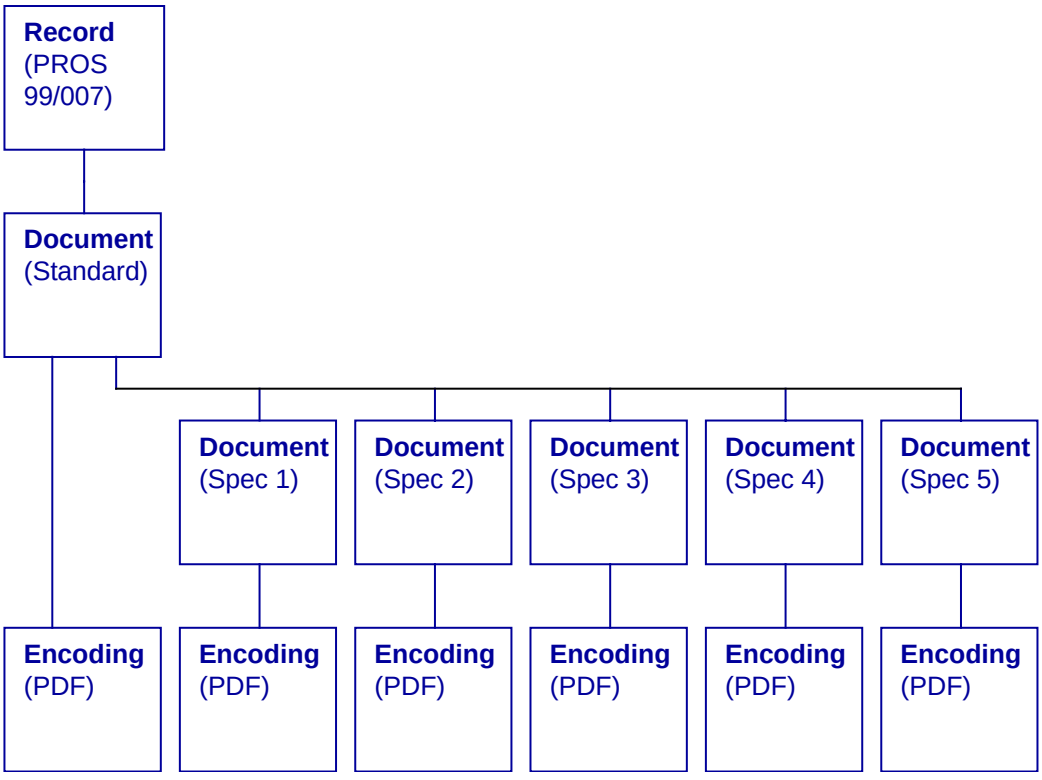


Figure 5. A record of PROS 99/007 with the five Specifications structured below the Standard.

6.1.2 Encodings

Encodings are different representations of the same document. For example, a VEO can contain PDF and Word versions of the same document.

The use of multiple encodings allows users to provide additional functionality for future users of the record. For example, a PDF representation of a report has to be included (as PDF is the long-term preservation format), but it may be useful to include a Word representation. The Word version may not be accessible for very long, but while it is accessible it would be possible to use the record as a template to create new records.

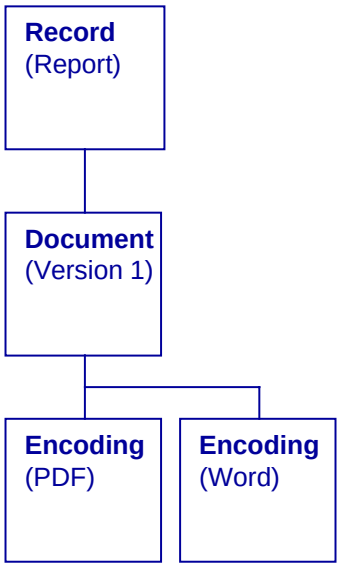


Figure 6. A record that has one document with two encodings. Both encodings represent the same content, but use different physical representations. In this case, one representation is PDF (the long-term preservation format) and the other is Word.

6.2 Record, Document, and Encoding metadata

Records, documents, and encodings can contain descriptive metadata.

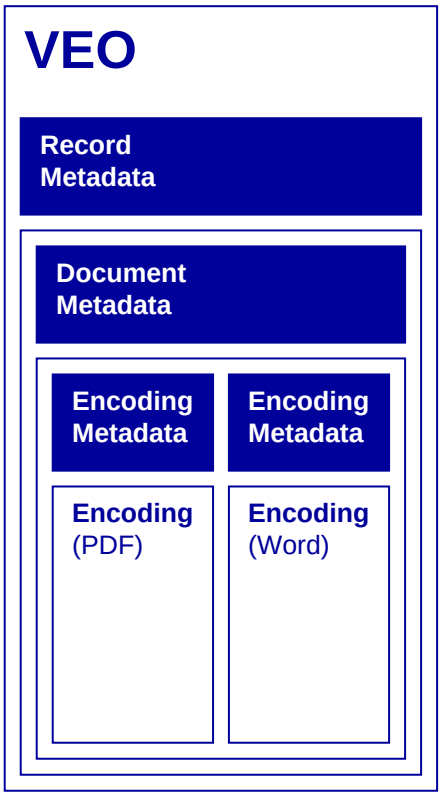


Figure 7. Record, document, and encoding metadata in a VEO.

Record-level metadata is intended to describe the record as a whole, including a description of the record and its context. Record-level metadata is based on the NAA metadata standard, with a few minor additions. The NAA metadata standard is based upon the AGLS [AGLS] metadata, which in turn is based on the Dublin Core metadata. Thus there is a strong family resemblance between Dublin Core, AGLS, NAA metadata, and the VERS record-level metadata.

Document-level metadata describes a document within a record. The description is in relation to the other documents within the record. Document-level metadata is also based on the high-level NAA metadata classifications.

Encoding-level metadata describes the encoding. The description is primarily technical, and has the goal of instructing future users (and computer systems) how to extract the document and render it.

7 References

[AGLS] AGLS Metadata Element Set, Part 2: Usage Guide, Version 1.3, National Archives of Australia, 2002, ISBN 0642 34491 4 (http://www.naa.gov.au/recordkeeping/gov_online/agls/metadata_element_set.html visited 29 May 2003).

[NAA] Recordkeeping Metadata Standard for Commonwealth Agencies, Version 1.0, National Archives of Australia, May 1999, ISBN 0 642 34407 8 (http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm visited 29 May 2003).

[PITT] Functional Requirements for Evidence in Recordkeeping, University of Pittsburgh, School of Information Sciences. Note that the original Web site documenting the results of this project was accidentally destroyed, but a copy can be found on the Internet Archive http://web.archive.org/web/20000818163633/www.sis.pitt.edu/~nhprc/ visited 15 May 2003.

[PROV1] Keeping Electronic Records Forever; Records Management Vision Development, Public Records Office Victoria, 1996, http://www.prov.vic.gov.au/vers/published/kerf.htm visited 3 July 2003.

[PROV2] Victorian Electronic Records Strategy, Final Report, Public Records Office Victoria, 1998, http://www.prov.vic.gov.au/vers/published/final.htm visited 3 July 2003.

[PROV3] Management of Electronic Records, PROS 99/007, Public Records Office Victoria, 2000, http://www.prov.vic.gov.au/vers/standards/pros9907.htm visited 3 July 2003.