Electronic Records – Problem Solved?: the Victorian Electronic Records Strategy and the future of electronic record keeping in Victoria

Ross Gibbs
Director
Public Record Office Victoria
ross.gibbs@dpc.vic.gov.au

Justine Heazlewood
Business Development Manager
Victorian Electronic Records Strategy
Public Record Office Victoria
justine.heazlewood@dpc.vic.gov.au

In 1998 Public Record Office Victoria (PROV) initiated a project to examine the capture and long term preservation of the electronic records of the Victorian Government. The project team built a system which successfully demonstrated that it is possible to capture electronic records with existing technology and from existing systems in such a way that electronic records can be preserved in the long term. Further work was begun in 1999 in the Victorian Department of Infrastructure where a project to implement a Victorian Electronic Records Strategy (VERS) compliant system is now well underway. In addition this year the PROV will release an electronic record keeping standard which will apply to all Victorian government agencies. These initiatives will change the way the corporate memory of Victoria and will enable future generations to access their history.

The electronic records 'problem'

Archival and record keeping systems to manage paper records have been part of government for centuries and have been developed to a sophisticated degree. Many Australian governments are working towards a situation where citizens transact business with government online. Much government business is already conducted, at least partially, electronically. This is especially true when one considers the growing use of email as an industry tool and the use of the internet, intranets and extranets, not only to communicate information to staff and clients, but also as a way of conducting business.

There are a number of reasons why electronic government activity should be recorded electronically rather than on paper.

- *The evidential status of a paper form of an electronic record is unclear.* An electronic record documenting a transaction which has been carried out electronically, may lose evidential weight if copied to paper, as it may not represent the 'original' record.
- *The cost of storing electronic records is much lower than that of storing paper records.* The decreasing cost of electronic storage combined with the increasing cost for paper storage means that it is no longer cost-effective to store all records in paper form.
- *A paper record of an electronic transaction may not capture all aspects of that transaction.* Printing electronic records to paper may not fully document the context of those records.
- *The cost of finding records is much lower and searching capability is greatly extended in an electronic environment.* Unless paper records are well managed and well documented, they can be difficult to find. An electronic environment allows more sophisticated ways of accessing and retrieving records.

The creation and management of records produced in an environment where desktop computers have replaced pen and paper has been of concern to both records managers and archivists for some time. There are a number of reasons for such widespread concern.

- *Document formats change and become unreadable*. For example, it is nearly certain that Microsoft Word documents that are created today will not be readable in 100 years time.
- *Electronic objects can be subject to undetectable change*. Unless precautions are taken, it is possible for an electronic file to be altered without any way of detecting that a change has occurred. The evidentiary status of that record may be compromised because it can be undetectably modified.
- *Electronic records may not be captured because most record capture processes are paper based*. Increasingly email, e-commerce, and electronic documents are being used to conduct business, but the processes of record keeping, developed over a long period of time, are paper oriented. To date much record keeping practice has required that electronic records are printed and then incorporated into the existing paper based record keeping system. Unless a conscious decision to print the electronic record is made, the record will not be kept.
- *The context of an electronic record, and its relation to other records, can easily be lost*. In an electronic environment the context of record creation can easily be lost if it is not documented at the time of record creation. For example, the time that a document is e-mailed to a client may be crucial, but this information is not usually recorded as part of the document.

- *Capturing context can be expensive*. Context is crucial to the understanding of the record in the future, but it may be too expensive to realistically expect to capture all the appropriate contextual information manually or to add contextual information to the record at a later stage.
- *Existing systems for managing electronic documents and records are not designed as archive systems*. An archive system preserves the content, structure, context, and evidential integrity of the record for as long as the record is required. Existing electronic document and records management systems do not provide the functionality required in order to archive records for long term accessibility.

Existing Solutions

Archivists wish to ensure that records which document government policy, individuals' rights and entitlements, and other classes of records identified as being of permanent value, can be managed in a way that ensures their continuing existence and accessibility to future generations. Archives and archivists (and indeed libraries and librarians) have been dealing with the 'problem' of electronic records in different ways.

When Public Record Office Victoria (PROV) began researching this problem in the early 1990s there were a number of 'solutions' put forward by various research institutions and archives. These were

- *Print to paper*: This is the current *ad hoc* solution. It may have unforseen legal implications in that a paper version of an electronic record may not fit the "best evidence" rule under Victoria's current evidence act. What, in fact, is the 'original' in the case of email? How do you guarantee authenticity when most email programs allow you to edit emails after they have been sent and received?
- *Technology museums*: This involves the literal warehousing of hardware and software in order to run old programs on the systems for which they were designed. This solution, while having a certain appeal, is probably impractical in the medium to long term.
- *Data migration*: This is advocated by many archival institutions. It involves agencies being made responsible for their records for all time (also known as distributed custody) and requires a commitment from agencies which is not assured (especially when responsibility may shift from the public to private sector or from agency to agency with sell-offs, mergers and restructuring). It also involves an increased cost to the agencies concerned.

PROV Response

Public Record Office Victoria (PROV) has initiated a series of projects designed to provide it with a new strategy for dealing with electronic records. In 1996 PROV was provided with funding from Victoria's Microeconomic Reform Fund to address issues of electronic records management and archiving. The project culminated in a report called *Keeping Electronic Records Forever*.[1]

---

[1]   *Keeping Electronic Records Forever: Records Management Vision Development,* Public Record Office and Ernst & Young, 1996. Available at http://www.prov.vic.gov.au/vers/kerf.htm

As the project was designed to establish a conceptual answer to a technological problem, the report developed a business case that explained the recommended solution in conceptual terms and identified the key issues upon which it impacted.

The project team concluded that there was a fundamental need for a 'static' *record* that was inviolable and which satisfied the evidentiary requirements of courts and government. It further concluded that agencies required 'static' *records* to which they could refer for their own operational needs, but also the full functionality that came with computer software applications, in order to create, modify and manipulate *documents.*

Victorian Electronic Records Strategy Project

In 1998 PROV began the Victorian Electronic Records Strategy project to implement the recommendations made in *Keeping Electronic Records Forever*. The project was run by the PROV in conjunction with the CSIRO and Ernst & Young. The project developed a prototype system to demonstrate that it was possible to capture records from existing business systems in such a way that they would be accessible in the long term.[2] The project team undertook as series of investigations so that they could understand the government processes which led to records creation and how records were used and managed in the same context. The archival processes of PROV were also investigated. Using this information, they pinpointed the information required to be captured when creating an electronic record.

Keeping Electronic Records Forever advocated that instead of taking a *system* oriented approach to electronic records, a *data* driven approach was more appropriate, as the records would outlast any system developed to manage them. This led to a recommendation for the adoption of a particular long term electronic record format (see below).

The second key point made in *Keeping Electronic Records Forever* was that electronic records should be captured at the time of creation. There are several strong reasons for this requirement. Firstly, the record is more reliable as evidence if captured at time of creation. Secondly, there is more chance that the record will in fact be captured if it is done immediately. Thirdly, information capture at the time the transaction is undertaken is both cheaper and more reliable than *post hoc* data entry. The project team also advocated that the capture of contextual information (metadata) should be as automatic as possible.

The VERS project was successfully completed in December 1998. PROV released the *Final Report* of the project in April 1999.

The VERS Demonstrator System

As part of the project, a demonstrator system was built to illustrate the principles outlined above. This demonstrator consisted of three major components:

- *Record capture*. This component simulated the desktop environment of a government agency. Record capture was implemented by integrating different desktop applications in representative workflows and capturing records into the Long Term Electronic Record format (see below).
- *Repository*. This component managed the archived records, including tracking the location of records, sentencing, and destruction.

---

[2] The project team took, as a putative figure, the period of 100 years to represent the 'long term.'

- *Record Discovery.* This component allowed users to search for and display archived records. The functions included in this component were the building of indexes from information contained in the records, searching for records, and the display of retrieved records as Web documents.
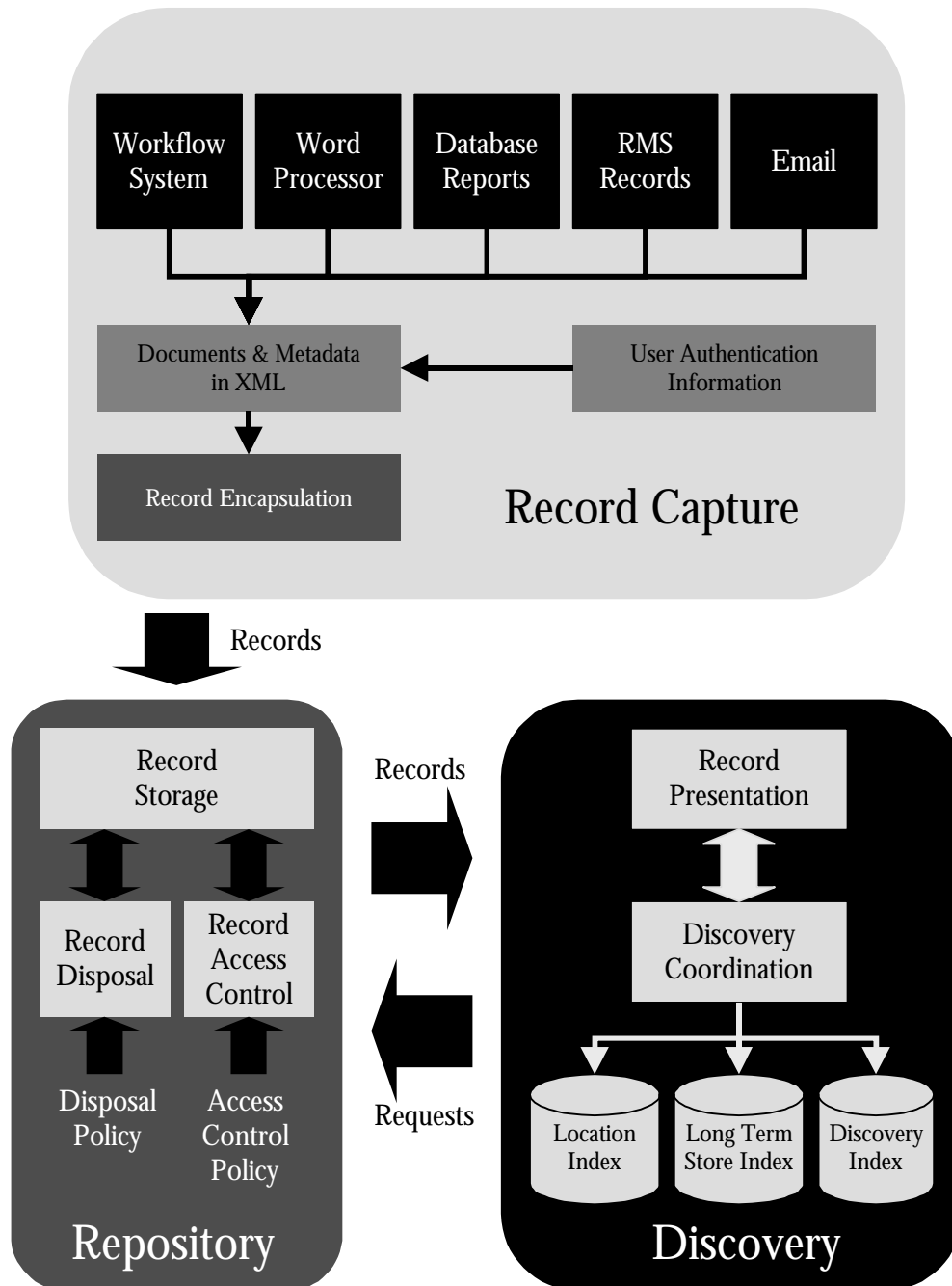
Figure 1 Diagram

```
┌─────────────────────────────────────────────────────────────┐
│  Workflow    Word        Database     RMS                    │
│  System      Processor   Reports      Records     Email      │
│                                                              │
│     Documents & Metadata        User Authentication          │
│     in XML                      Information                  │
│                                                              │
│     Record Encapsulation        Record Capture               │
└─────────────────────────────────────────────────────────────┘
```

Figure 1: Demonstrator System Design

Record Capture

From an archival perspective, it is important that both the content and structure of records are accurately captured. The project determined that the captured record should be identical in appearance to the original document as it was viewed by the creator of the record. This requirement restricts the technologies that can be used to capture documents. With HTML (Hypertext Markup Language, used for World Wide Web documents), for instance, the

appearance of the document depends on the settings of the user's browser and the resulting record may bear no resemblance to the appearance of the original document.

The VERS project determined that record capture systems could automatically generate much of the metadata required to document the context of the record thereby relinquishing the need for manual entry on the part of the record creator or the records manager. Metadata which depends on the record capture system can be automatically generated by that system. For example, all records produced by a particular process will have the same technical description. Metadata can be derived from the document itself, although this can be application dependent. For example, email headers contain information about the sender, the recipient, the time and date of sending, and the subject of the email. Metadata can also be derived from the computing environment. In particular, the record's creator and its time of creation (essential for the evidentiary status of the record) can be obtained from a smart card or the user login. Finally, a record generated from a programmed workflow automatically has a context (that is, a relationship with other records in the workflow). Metadata entered at one stage in the workflow can be carried along with the workflow and added to later records.

Repository
Archiving is the process of managing records over a long period and includes storing, preserving, and providing access to records. There are a number of issues which arise when these functions occur in an electronic environment. The most pressing issue is technology obsolescence.

Computer technology rapidly becomes obsolete. Therefore it will be necessary to move records from one medium to another as storage technology changes. This process is often termed 'refreshing.' Refreshing may also be carried out to protect against record loss due to deteriorating media (for example a tape wearing out), or to make a physical copy of the record (for off site storage for instance). The process of refreshing can be made entirely automatic.

Discovery
Being able to follow contextual relationships (like the link between incoming correspondence and outgoing replies) provides a powerful mechanism for finding information. Most of the contextual information used in the VERS demonstrator system was captured when the record was created and was stored as metadata within the record. The basic contextual relationship used in the Victorian government is collecting related records into files. To aid in management of records, particularly at the PROV, files are assigned to series (collections of related files) and series to agencies.

Officers who access files on a day to day basis will understand the purpose of individual files within the agency context. To assist other users and researchers' understanding of the context of the records, archival finding aids are created to describe the history and function of the agency, as well as the purpose and organization of the files and series within the agency. Finding aids were able to be stored within the demonstrator VERS discovery system.

There are many other contextual relationships which can also be documented. *Ad hoc* relationships are essentially random links between related files or records and may be created for any purpose at any time. A typical ad hoc relationship is a linkage between versions of a record. *Ad hoc* relationships can be documented in a record's metadata but the user creating an *ad hoc* relationship must consciously document that relationship.

The VERS demonstrator made use of Web based technologies and delivered records via a Web browser. As the archived records were kept in a database, and had extensive metadata associated with them, it was possible to construct a much more powerful searching environment than with paper based records. The Discovery System allowed:

- *Searching via a finding aid and original indexes*. This reflected current practice within PROV, where records are found by searching the original agency indexes. PROV also develops finding aids to describe the purpose and structure of the agency and the files within the agency.
- *Searching the content of the metadata*. Searches were able to be performed on any combination of the metadata fields.
- *Searching the content of documents within the record*. The VERS demonstrator system stored the text for each record in the database so that it was possible, for example, to return all records that contained a specific word.
- *Searching on any combination* of the content of the documents and metadata.
- *Searching on record context*. The VERS demonstrator allowed records from a particular transaction to be linked together.
- *Searching on File context*. All records were associated with a File, and a group of records in the same File were able to be retrieved.

Electronic Records

Electronic records are simply the computerised versions of traditional paper records created and kept by agencies. Sources of electronic records range from desktop applications such as Word, Excel, and email, to corporate applications such as financial systems, HR systems and corporate databases. Typically records are evidence of government or organizational activities and include policy documents, memos and letters, and database reports. Theoretical archival science contains a number of formal definitions of 'electronic records' but, from a practical perspective, the project needed a definition which gave the widest possible scope to agencies and PROV to define records for their own purposes. A record, to an agency, is simply whatever information they need to preserve. Victorian agency decisions will always be directed by legislation, PROV standards and agency practice.

The project concentrated on the preservation of electronic documents. Documents comprise the majority of government records, and most documents are created electronically on desktop computers. The project also decided that focusing on a manageable subset of records would be more fruitful than attempting to cover all types of electronic records.

Long Term Electronic Records

Electronic records which are able to be accessed in the long term need to meet the following criteria:

- *Long life*. Records must have an indefinite life. That is, the contents of a record must be capable of being viewed forever as the users originally saw them. Therefore the records must be in a form that can be physically preserved (for example easily copied from one media to another). It is also essential to preserve the indices and context of the record as to preserve the content itself so that records can be found. The context should be sufficient that the record is able to be understood.

- *Evidence*. A major reason for preserving records is to be able to legally prove what actions were taken and why they were taken. Electronic records must consequently be admissible as evidence and given due weight in a court of law. In practice this necessitates the ability to prove who created the record, when it was created, and that the record has not been subsequently altered.
- *Disposal*. Not all records need be preserved forever. Some records, indeed, must be destroyed after a period. Long Term Electronic Records must be able to be sentenced and destroyed if necessary.
- *Augmentation*. Not all information about a record is known when the record is created. A record may, for example, need to be refiled. It is necessary to be able to augment or change the information associated with a record without disturbing the evidentiary integrity of the record.

Long Term Electronic Record Format

A good long term electronic record format should be:
- *Self documenting*. It must be possible to interpret and understand the information in the record, at least at a primitive level, without reference to external documentation (which might have been lost). To this end the recommended structure
  - Is based on ASCII text. This means that the structure of the record can be viewed using the most primitive of computer tools.
  - Contains short textual descriptions of any more complex encoding of information.
- *Self contained*. The electronic record structure must contain all information about a record. It is far easier and more reliable to manage the information associated with a record if it is stored in one place rather than in components which are stored separately.
- *Extensible*. It is simple to extend the structure to add new metadata or new record types without affecting interoperability of the recommended basic structure.

The VERS Long Term Electronic Record format is basically XML (eXtensible Markup Language). XML is a text based mark-up language and is easily extensible (unlike HTML) and relatively simple. A number of other document formats can also be used within this XML structure. The VERS demonstrator encoded document content using Adobe System's Portable Document Format (PDF) Version 1.2.

The standard record structure is shown in Figure 2. A generic VERS record may contain many different types of object; in this case, it contains a record. A record contains metadata general to the entire record and one or more documents. A document contains metadata specific to itself and one or more encodings of the document. An encoding contains a representation of the actual content and metadata that describes how the content was encoded for inclusion in the record.

Each signature block contains the result of signing of the record. The signature validates the preamble and the object content. The signature blocks are included for evidential reasons. It is necessary to be able to show who created the record, when it was created, and that the record has not been subsequently altered. This is required not only to guard against forgery and alteration by the creators of documents, but also to guard against forgery and alteration by the administrators of the Archive system itself.
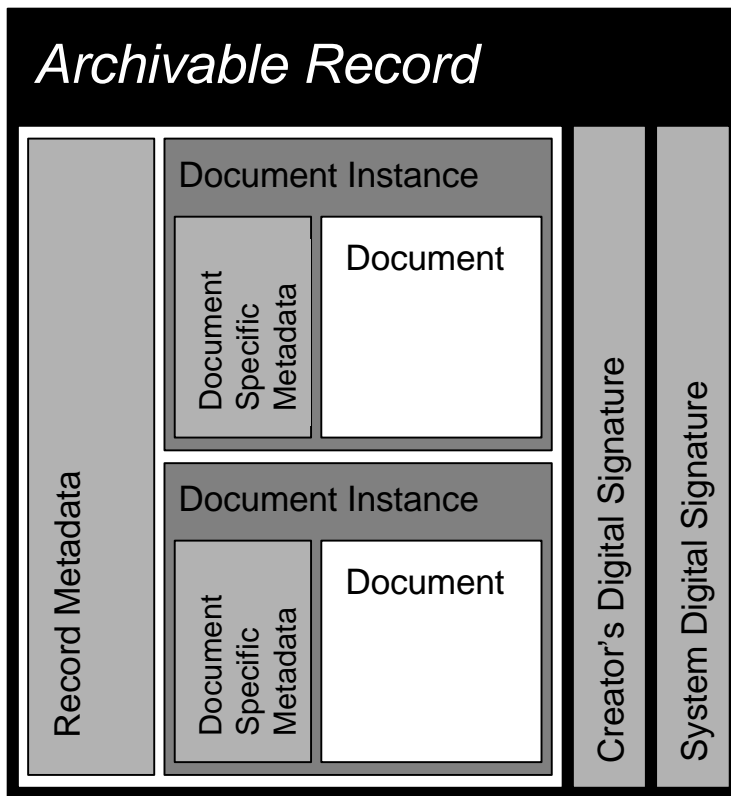
Figure 2: Standard Long Term Electronic Record Structure

In the VERS model, documents are represented using PDF, although other document formats representing the same document may be included in the record. The primary selection criteria for a document format was confidence that, for the foreseeable future, it would be possible to write a viewer for the document from publicly available information. Word file format, for instance, would not be an appropriate format, as the description of this format has not been published. The PDF standard has been published and is freely available.

PDF is also flexible. PDF can be generated from any application that can generate Postscript (the standard printing language); thus anything that can be printed can be represented in PDF.

Augmenting Records

The digital signatures that secure a VERS record also prevent any modification of the record. In some circumstances, however, it might be necessary to modify the metadata associated with a record. For example, the need may arise to refile a record, to add additional descriptive information, or to make a new linkage to another record. To allow the metadata to be modified without disturbing the evidentiary integrity, the VERS approach allows an entire record to be included as the content within a new record. This layering of record metadata is referred to as creating 'Onion Records' (see Figure 3). The content in the case of an Onion Record is a complete record.

Where is VERS Going?

Standard on Electronic Records Management

PROV legislation provides for the mandating of Standards for records management for the Victorian government. To this effect PROV has produced an exposure draft of its *Standard on the Management of Electronic Records* (PROS 99/007) which includes specifications on the

- *Generic VERS record structure*: an understanding of the basic record structure – as outlined above.
- *VERS system requirements:* What sort of functions does an electronic record keeping system need to perform?
- *Standard Electronic Record Format*: including a technical description of the standards used (XML, PDF, etc).
- *VERS metadata scheme*: this scheme has a direct correlation to what is in the new National Archives of Australia's Recordkeeping Standard for Commonwealth Agencies.

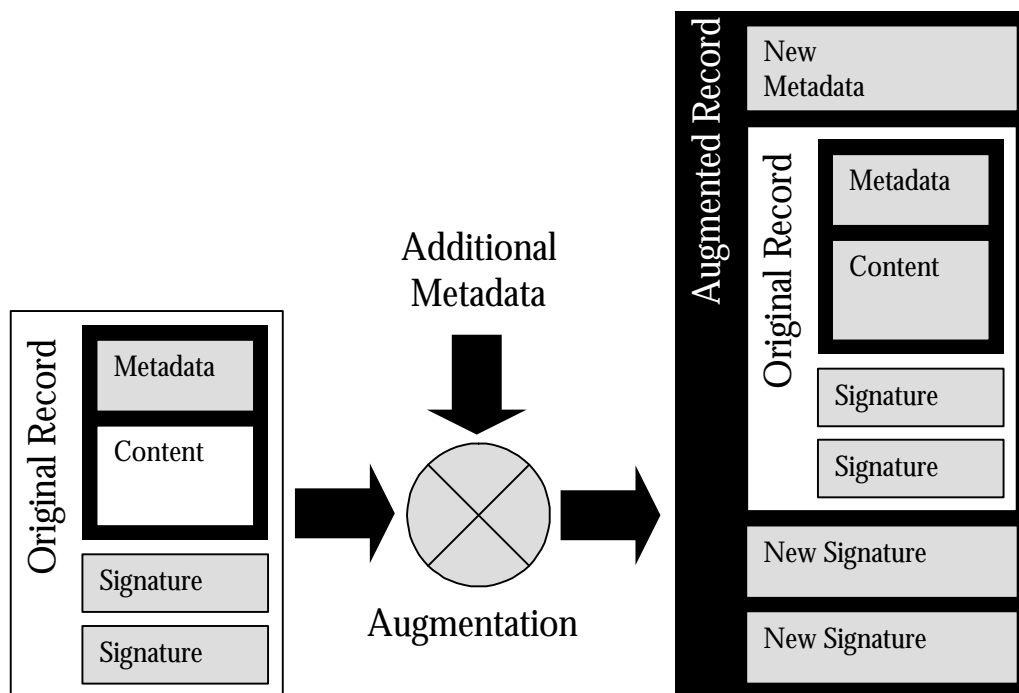This Standard will be launched to the whole of Victorian government in March 2000.

Figure 3: Augmented Record

Department of Infrastructure Implementation

The Victorian government has continued to supported the Strategy with $4.8 million to undertake a staged implementation of a VERS compliant system within the Department of Infrastructure. This implementation will provide a model for the rest of the Victorian government a commercial prototype as an example for vendors.

Conclusion

The role of the records manager becomes increasingly important in an electronic environment. The context of individual records and files can be captured using the VERS model, but more global information about the organisation as a whole could easily be lost. Records managers are needed to ensure that information about the *way* that an organisation keeps records, as well as the records about the organisation as a whole, survive the move to a digital world.

The VERS project has successfully demonstrated that the capture and long term preservation of electronic records is possible now. The VERS project offers a solution to the electronic records 'problem' which has plagued the record keeping profession for some time. Further work needs to be undertaken to deal with very complex electronic records, but the VERS approach gives us the ability to tackle the clear majority of the electronic records being created now.

Full details of the VERS project can be found in the Victorian Electronic Records Strategy Final Report available from Public Record Office Victoria or at http://www.prov.vic.gov.au/vers/.