

# Electronic Records: Challenges Overcome, and Issues to Come

Andrew Waugh



# A journey of a thousand *li* starts beneath one's feet

- Lao Tzu (Archivist at the Imperial Library, Zhou Dynasty court during the 6th century BC)

# Your mileage may differ

My observations are based on what I have observed in Australia, and read (in English). But recordkeeping has a cultural aspect – what is hard in Australia might be easy in your culture.

**My views are not official views of the Public Record Office Victoria**

# My thesis

Digital recordkeeping is hard because we persist in trying to manage digital records using the same approach we successfully managed paper records

# Roadmap of presentation

- PROV's digital record journey so far
- Email as a recordkeeping challenge
- What is causing these recordkeeping challenges?
- Where to now?
- Conclusion: it is still a journey



# Public Record Office Victoria

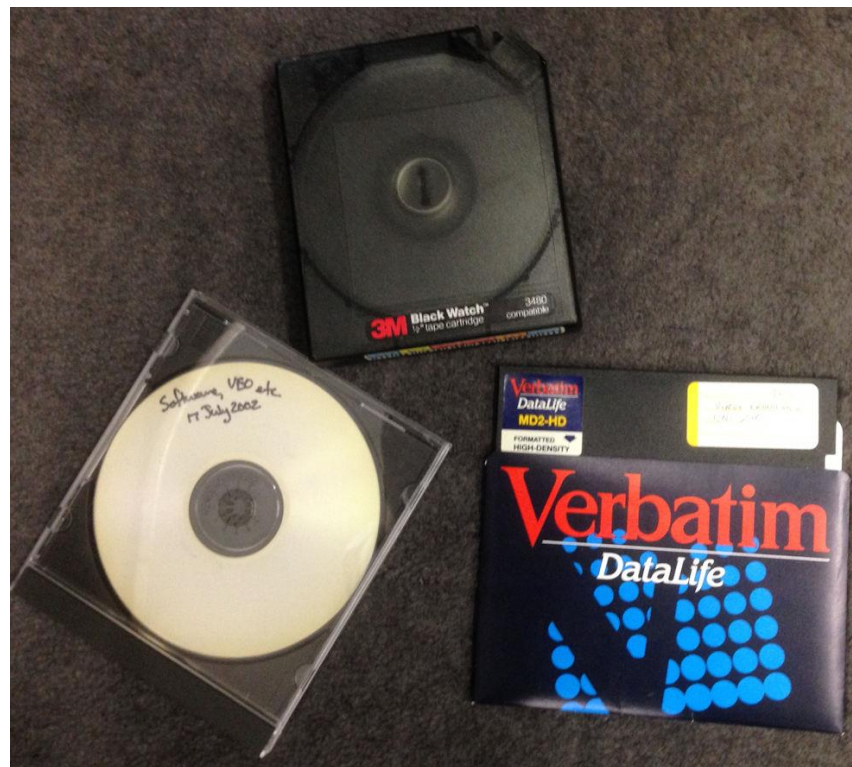


# In the beginning...

*Digital information lasts forever – or for five years, whichever comes first – Jeff Rothenberg (1995)*

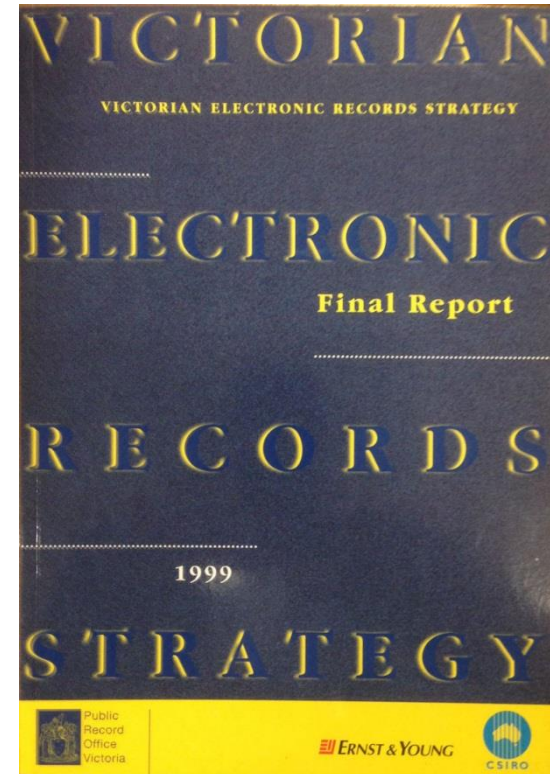
In the mid 1990s the archival community was beginning to worry about the preservation of digital objects

A lot of work was done by archivists on this problem



# Was there a way forward?

- In the late '90s PROV undertook research into digital preservation (with CSIRO and E&Y)
- Culminated in 1998 with the publication of the 'Victorian Electronic Records Strategy' final report (aka VERS)
- Builds on existing work – David Bearman & other Australian jurisdictions





# From a Strategy to a Standard

- Publication of a PROV standard “Management of Electronic Records” (PROS 99/007) in 1999
- Trial implementation in a government agency (VERS@DoI) in 2000-2002

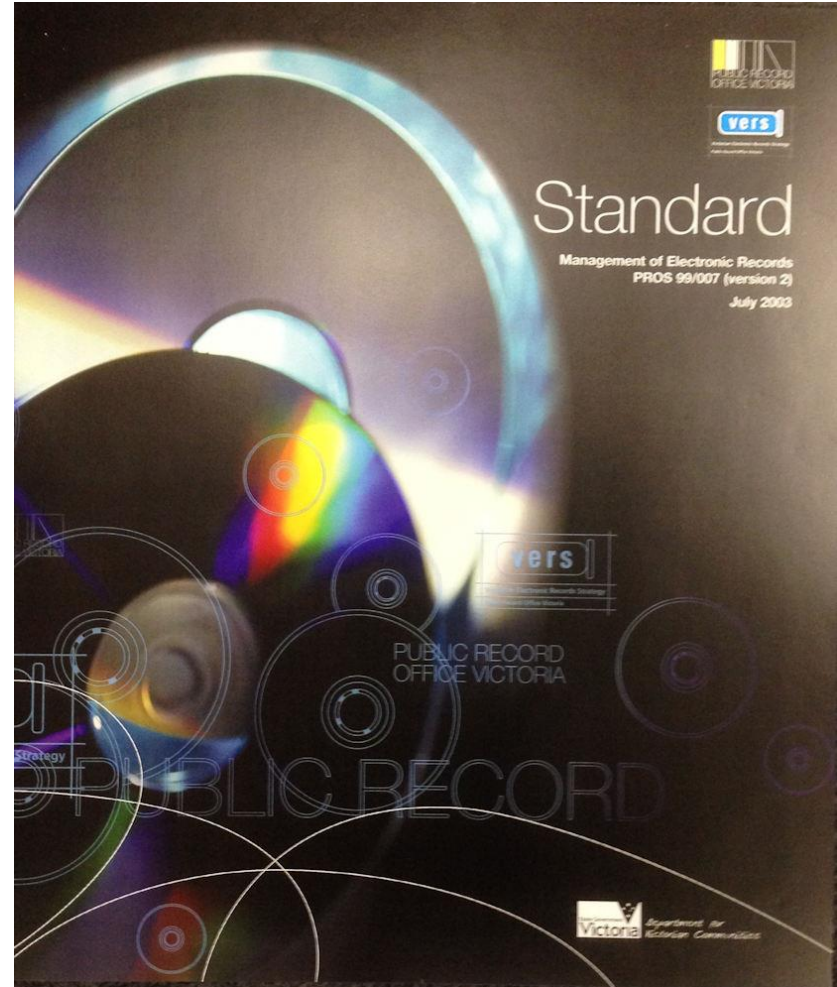
# Steady as she goes

## VERS Centre of Excellence in 2002

- Implement a digital archive (live in 2004)
- Second version of VERS (revisions, and more explanations)

## Production of tools, compliance regime, accepting records

- Steady stream of digital transfers



# VERS renewal

New standard in 2015  
(PROS 15/03) building  
on lessons learnt

- Including validation & construction tools

Building new Archival  
Business System

- Including a new digital archive that will accept VEOs to new standard

## Standard

PROS 15/03  
Standard for the encapsulation of digital information

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

## Specification

Constructing VERS Encapsulated Objects (VEOs)  
PROS 15/03 S1

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

## Specification

Adding metadata to VEOs  
PROS 15/03 S2

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

## Specification

PROS 15/03 S3  
Long term preservation formats

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

# VERS is...

A framework for capturing and managing digital records

- Provides a framework for advising agencies about digital recordkeeping issues
- Structures transfers to the archive
- Technical:
  - Capture standard metadata about records
  - Record content in a standard long term preservation format
  - Lock metadata and content into a single managed object
  - Digitally sign to detect corruption



# How naïve – a technical solution

Our recordkeeping journey so far...

- Technically we are confident we can preserve and make available digital records
- But our agencies still struggle with the actual creation and ongoing management of digital records
- (But not digital information)

Why?

# Email as a recordkeeping challenge

Email is a microcosm of the challenges facing recordkeeping in the digital age



# Secretary of State Clinton



Used personal email to transact business as US Secretary of State

- 100,000 pages of official records
- Records only recovered because of politics
- Example of BYOE (Bring your own environment)

# **Our investigators look at the records, *but the smoking gun is always in the email***

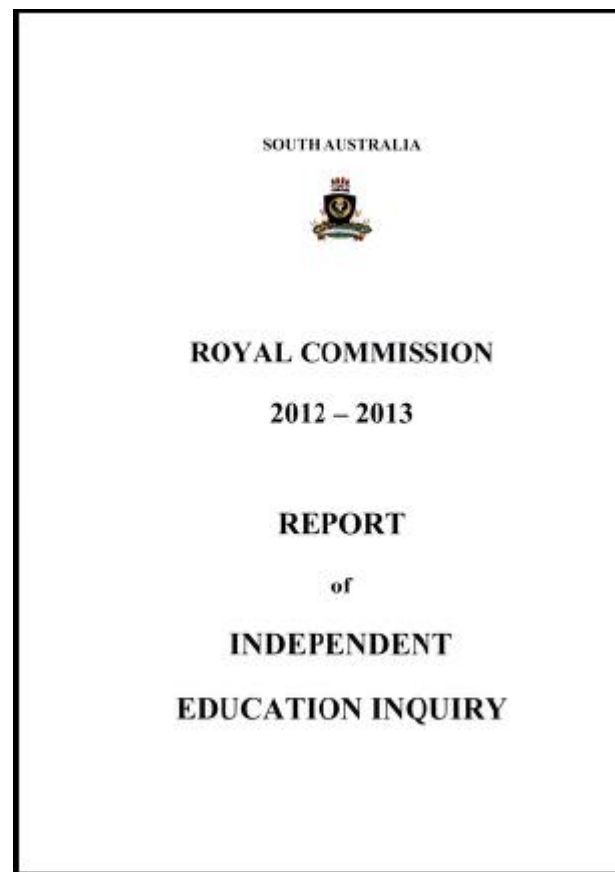
- Victorian Auditor General & Victorian Ombudsman to records managers c2010



# Debelle inquiry (1)

## Typical investigation

- Email trail shows what the participants did and said at the time
- Unconscious creation of records
- <http://www.saasso.asn.au/wp-content/uploads/2013/11/DebelleInquiry.pdf>



# Debelle inquiry (2)

But the downside of emails was...

- It was hard to find relevant emails
- The email trail was not complete due to uncontrolled deletions
- Uncontrolled deletion of emails embarrassed the relevant minister at the time – who was then the Premier



# Integration into ERMS

- In 2011 the US State Department created 61,156 record emails out of 1 billion emails sent
  - State has an email/EDRMS integration since 2009
  - OIG's main recommendations were more training
  - Email integration clearly doesn't work in many organisations
  - Very conscious creation of records



## Review of State Messaging and Archive Retrieval Toolset and Record Email

### DOMESTIC OPERATIONS AND SPECIAL REPORTS

**IMPORTANT NOTICE:** This report is intended solely for the official use of the Department of State or the Broadcasting Board of Governors, or any agency or organization receiving a copy directly from the Office of Inspector General. No secondary distribution may be made in whole or in part outside the Department of State or the Broadcasting Board of Governors, by them or by other agencies or organizations, without prior authorization by the Inspector General. Public availability of the document will be determined by the Inspector General under the U.S. Code, 5 U.S.C. 552. Improper disclosure of this report may result in criminal, civil, or administrative penalties.

SENSITIVE-BUT-UNCLASSIFIED

# \$970,252 for an FOI request

In 2012 Cenitex argued that it would take 24 years and \$970,272 to search 22 months of email to satisfy Freedom of Information requests

- The court accepted the argument
- Worse, it could actually be true
- Issue was the use of backup as an 'archive'





# Summary of issues

The records our stakeholders want are not in the recordkeeping systems

- Staff are working in ad hoc digital systems – often their own systems
- Records that are created are not captured into record systems
- This means that it is difficult/expensive/impossible to provide access, manage, or control destruction of the records
- But these systems provide rich sources of records that our stakeholders want

# What is causing these problems?

(For email and similar other digital records)



# What are records?

- Information created, received, and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business (ISO 15489.1-2002)

# Why do we keep records?

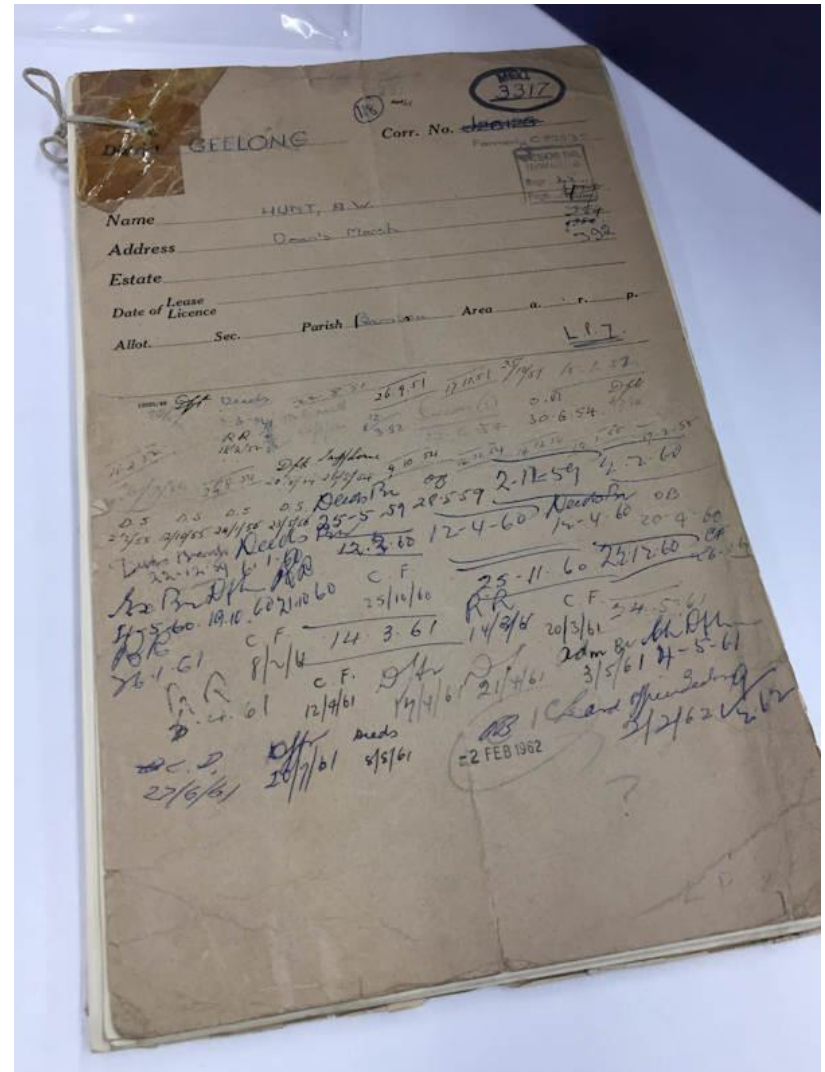
- To document people's entitlements
- To efficiently carry out work over time
- To protect our legal position
- To hold organisations and staff to account
- To provide society's memory
- To reuse valuable information





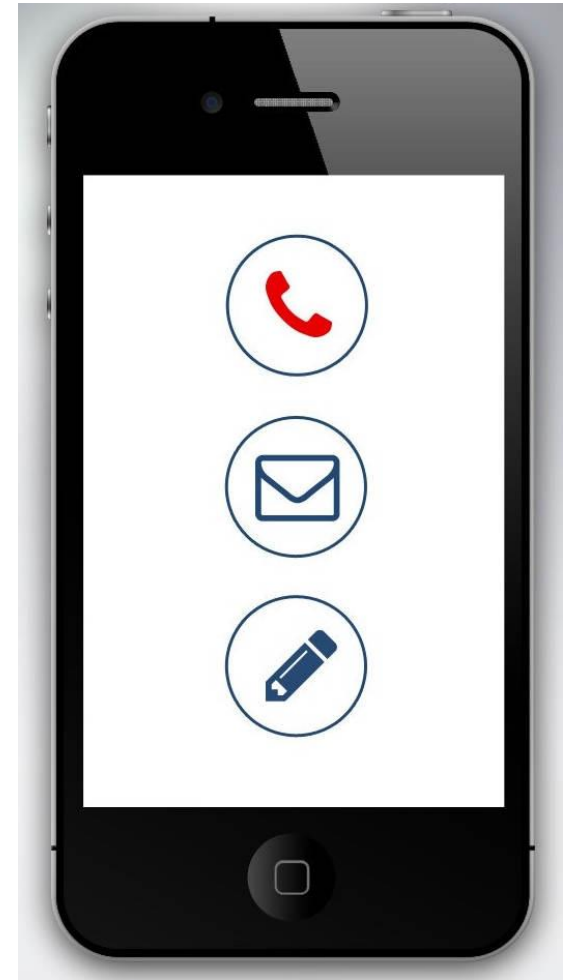
# Work is about collaboration

- In the paper world, the file was the collaborative mechanism
- Records were created as a side effect of collaboration
- This unselfconscious creation is the source of the records' reliability and authenticity
- (It was not perfect, of course)



# But collaboration has changed

- The modern workplace has many more means of collaboration (i.e. working)
  - Email
  - Shared drives
  - Business systems
  - Voicemail
  - Twitter
  - Facebook
- Work is largely done outside the formal 'record' systems



# Why is the smoking gun always in email?

Because email is where people are doing their work

- Email is generated as a side effect of doing business

Email captures transactions at a lower level

- Records are created of interactions that were previously unrecorded

# We have been side tracked by EDRMS

Modern EDRMS developed from systems for managing paper files

- Allow you to 'register' digital objects onto 'files'
- We recognise minor problems
  - Most people are resistant to expending the effort necessary to classify and capture the records
  - Made worse by clunky, hard to use integrations
  - Cloud based email adds a whole new dimension of barriers
- But the real problem is
  - Recordkeeping is separated from the doing of the work

# Records are where the work is

- We have tried to manage digital records in EDRMS with a paper model – the file and the filing system
- But work has moved to new systems – such as email
- These systems have not been built or designed with records in mind
- Why are we surprised that records are not in the recordkeeping systems?
- These changes in work practices are still happening – email is just the beginning



# So - where to now?

Using email to investigate challenges of moving to where the work is



# Capstone

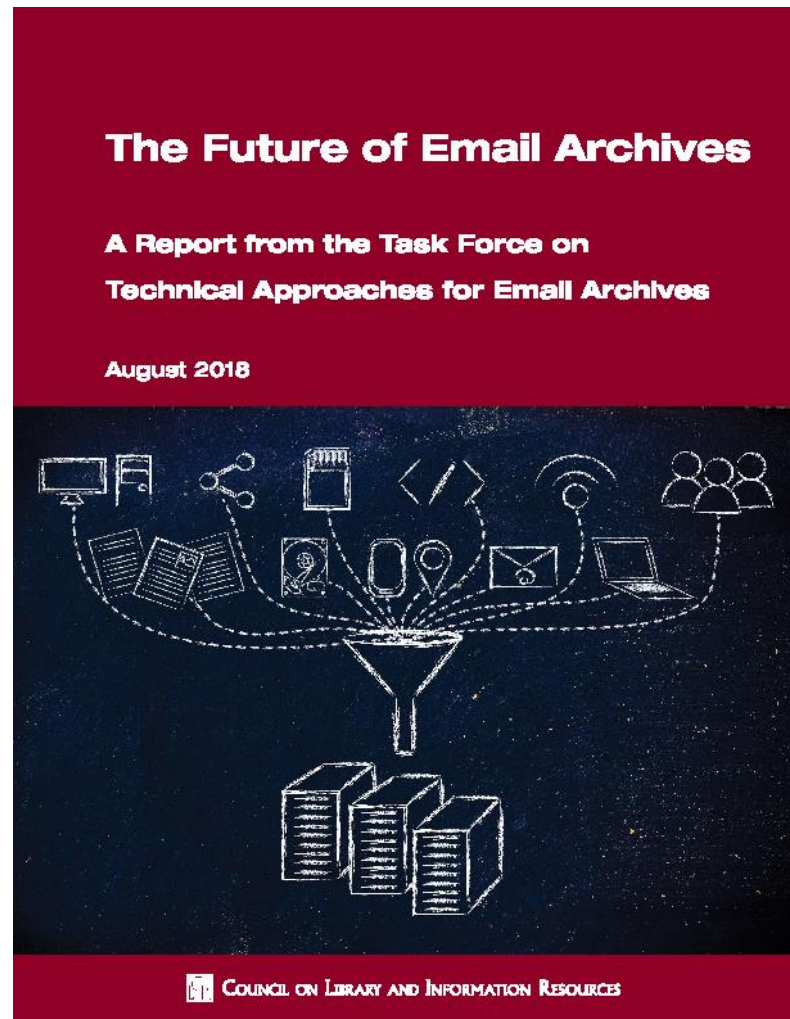
In 2013, NARA turned its advice on email management on its head

- Permanently keep all email of key staff
- Temporary retention of all email of all staff
- Qualifications – NARA still wants agencies to do the usual record management stuff (disposal, prevent modification or unauthorised deletion, access, appropriate metadata)
- <https://www.archives.gov/records-mgmt/email-management/capstone-training-and-resources.html>

# CLIR Report on email archiving

Exhaustive review of state of the art in archiving email

- Focused on archives, not operational records in agencies
- <https://www.clir.org/pubs/reports/pub175/>



# Pilot capture project

PROV used an eDiscovery tool to investigate email capture and filtering

- 1.5 Terrabytes from 1480 staff, 4.6 million emails with 33 million objects
- 43% emails removed in deduplication
- Appraisal tests
  - Negative (explicit tests for uninteresting domains) removed a further 7% at 100% accuracy
  - Positive (lists of interesting keywords) also removed 7%, but at 98% accuracy
- Tested analysis (security classification, encryption)

# Access becomes the key issue

The key weakness of bulk capture is supporting access

- No linkages with other collections of records
- Records are split between user mailboxes, and are necessarily organised even within the mailboxes
- No ability to control access (security & privacy)
- Open problem, but...
- We should look at work being done in the big data community on visualising large data sets. Exploit the data and metadata in email – build on its strengths



# Visualisation possibilities

Access is a visualisation problem – what views of a collection can be provided to facilitate different types of access?

- Visualisation can be used in agencies as well as archives
- Visualisation techniques will improve (be dynamic not autocategorisation)
- Visualisation techniques can be used positively and negatively
  - Positively to find and organise
  - Negatively to hide the ephemeral and private
  - Ultimately dispose of uninteresting emails???

# Conclusion on email capture

## Our current thinking?

- EDRMS integration only if there is a good reason
- Capture all email and keep for a period (possibly permanently)
- Email must be retrieved in a usable form from system
- This collection forms foundation for future work
- Explore visualisation/data mining techniques to dynamically organise email to facilitate access (and suppress records)
- Techniques (access) will get better and better as other disciplines invent new approaches

# Key Messages

- We are still on a journey
- We now realise that
  - the key challenges of digital records are about people, they are not technical
  - The best records are created as a side effect of carrying out work
  - These records are usually found in the systems that created them, not in recordkeeping systems
- Challenges are:
  - Bringing the records under management
  - Providing (and controlling) access (telling the story)z

Thank you...  
...any questions?