

Email

Where to from here for PROV?

Andrew Waugh, David Brown & Howard Quenault

Secretary of State Clinton

Exclusively used personal email to transact business as US Secretary of State

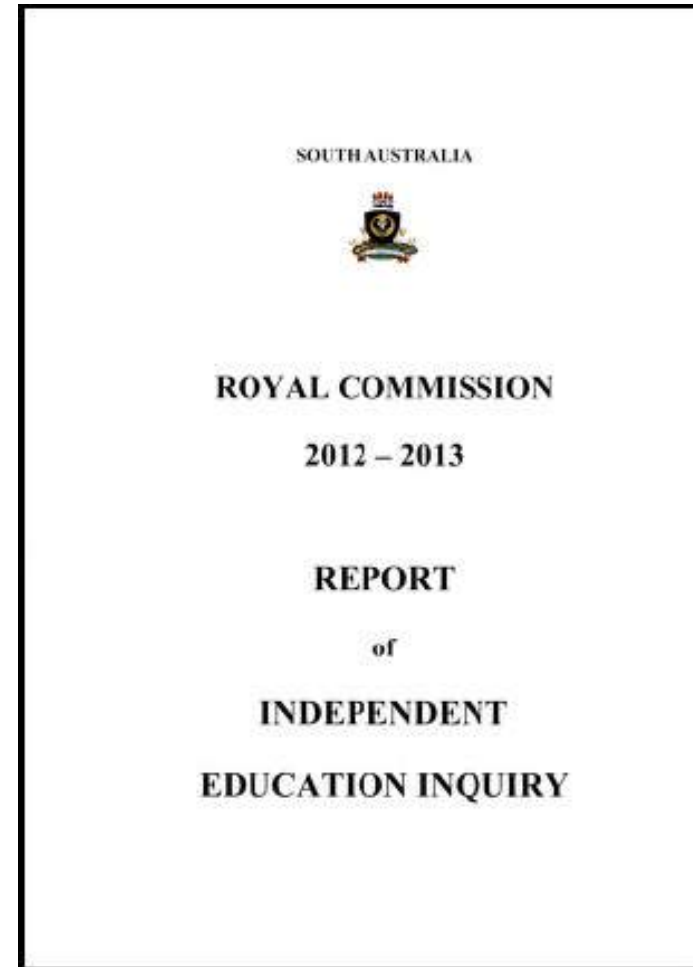
- 100,000 pages of official records
- Records only recovered because of politics
- Example of BYOE (Bring your own environment)



Debelle inquiry (1)

Records are important, but the smoking gun is always in the email

- Email trail shows what the participants did
- Because email is created as a by-product of doing work
- This is unselfconscious creation of records



Debelle inquiry (2)

But who was told and when?

- 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 is now 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



\$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 FOI requests

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



So, how does email stack up?

To answer this question, consider why 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



Positives and negatives of email

Positives

- Generated as a side effect of doing business
- So the record is more likely to exist and be accurate (good for entitlements, work, legal, governance, memory)

Negatives

- Deletion is not controlled
- Which means the record may not exist (but multiple copies usually exist)
- Access is very problematic (record is held in individual mailbox, and can be hard to find even there)
- But we don't exploit features of email

Management by EDRMS

Classic approach by Records Managers

- Still recommended today (particularly in the US)
- Many people are resistant to expending the effort necessary to classify and capture their records
- Made worse by clunky, hard to use integrations
- Cloud based email adds a whole new dimension of barriers
- There is a fundamental archival problem – conscious creation of records

Capstone

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

- Permanently keep all email of key staff (but which key staff?)
- Temporary retention of all email of all staff
- Qualifications – still need to do the usual record management stuff (disposal, prevent modification or unauthorised deletion, access, appropriate metadata)

Possible PROV approach

- Not fixed policy yet, but ideas
- Focus on foundations
 - We don't know what technology will bring
- Limit our call on agency and staff capabilities
 - Use our limited influence wisely



Basis... Don't mess with the email system

Capture and manage email in the email system, export it to an archival system, and retain it all for a period of time

- Necessary for existing email collections
- No change in how users use email
- Records are created as a side effect of doing business
- Matches our approach with other non-traditional record system (e.g. business systems)
- An aspect of Capstone
(but NOT who gets captured permanently)

Spend our effort on battles we can win

We have little influence on the systems agencies deploy, so we must use that influence wisely

- Critical choice about what functions we push for in operational systems
- We're still discussing this!
- Must: Must be able to extract records from email systems in a usable form
- Like to have: Disposal and integrity control
- Questionable: Classification
- Someone else's problem: Security and privacy

Access becomes the key issue

The key weakness of Capstone is supporting access

- 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 and data mining large data sets. Exploit the data and metadata in email – build on email's strengths

Visualisation possibilities

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

- Visualisation/data mining techniques
 - can be used in agencies as well as archives
 - will improve (never do autocategorisation once, now)
 - can be used positively to find and organise
 - can be used negatively to hide the ephemeral and private
 - could be used to dispose of uninteresting emails

Conclusion

A work in progress, but 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 (other functions still debatable)
- Resulting 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

Questions

#RMN2015
@PROV_RKS

