Defusing the ticking time-bomb

How should Records Management deal with email archiving?

Chris Gousmett, Corporate Information Manager, HCC
ALGIM Conference, Wellington, 22-23 May 2017
The story begins here…

**Police rocked by porn scandal**

By Ainsley Thomson, Louisa Cleave, Helen Tunnah

8:37 AM Friday Apr 22, 2005

The pornography scandal engulfing the police was uncovered during a wide-ranging probe of police culture sparked by a string of damaging controversies last year.
The problem

• Email “archives” were implemented across many government and local government agencies after inappropriate cartoons were found on Police computers in 2005.

• This addressed an immediate problem: how to track what emails were being sent and received, and how to enable auditing of content.

• It also created a new problem: what to do with all the emails which are being retained?
Inaccurate justification

- Misconception that “records management” means “keep everything forever”
- “There is a legal requirement to capture all email in and out of Hutt City Council” – Business Case for Email Archive 2006
- Implementing an Email archive would then ensure “full compliance with the Public Records Act 2005 and Electronic Transactions Act 2002.”
This was challenged at the time

• Email archiving is not active management but passive management
• It captures many emails of low information value
• Not a controlled and managed environment
• Emails accessible only to IT and each user can access their email – does not meet requirement for accessibility
• Cannot apply disposal criteria
• Emails of archival value cannot be accessed by public
The solution

Ignore it and it will go away...
Meanwhile, in the background... 

- Thousands of emails continue to be created daily
- They are all captured by the archiving product
- No visibility on what is happening
- Not included in record-keeping processes
Then this happened...

- January 2017 - Archives NZ issued a survey to central and local government asking what information was held
- A request to IT for some statistics on the Email archive disclosed that we had over 30 million emails totalling 2 Tb
- This unmanaged and unmanageable stash of information suddenly gained visibility with senior management as a risk
The elephant in the room

“I suppose I’ll be the one to mention the elephant in the room.”

Many of these emails are records of business activity
They need to be managed as records
They are not held in any other record-keeping systems
Naming the Elephant

• How do we address the problem?
• Firstly, get some visibility of the systems being used
• Get access to the email archive for records staff
• Find out what tools are available to access the systems to enable some analysis
• Once you have an analysis of what is in them, what can you do to manage them?
• How do you set up a sustainable process for the future management of emails?
Are they Records?

• If they document the business of the organisation, yes they are records.
• Many of them may be transitory records which do not need to be retained for long
• Others may be critical evidence of business activity which is not captured in any record-keeping system
• They are subject to the Public Records Act, LGOIMA and Privacy Act.
Risks in Email Archiving

• Public Records Act
  – Records to be retained in accessible form as long as necessary – not generally accessible
  – Authorised disposal when no longer required – not easily done
• LGOIMA
  – Information needs to be made available on request - searches on email archive not routinely done – could miss important information
• Privacy Act
  – If someone requests access to all information we hold on them, could we comply?  [Principle 6]
  – Information is not to be kept longer than necessary [Principle 9] – Information not to be used for purposes other than the original purpose for collection [Principle 10] – but do we know what that was?
• LIM reports – do emails contain information we need to supply?
The current situation

- Over 30 million emails from 1997 onwards (back-captured in 2010 when current system was put in)
- Averaging 200,000 per month over the last 10 years
- Frequent need to access emails to plug gaps in the record
- No systematic method of dealing with the valuable records mixed in with personal stuff and plain junk
- Only IT staff could access the email archive
The current situation
The current situation

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email Message</td>
<td>25,378,371</td>
</tr>
<tr>
<td>Document</td>
<td>839</td>
</tr>
<tr>
<td>Contact</td>
<td>636,694</td>
</tr>
<tr>
<td>Task / To Do</td>
<td>42,697</td>
</tr>
<tr>
<td>Task Request</td>
<td>397</td>
</tr>
<tr>
<td>Journal / Notebook Entry</td>
<td>18,685</td>
</tr>
<tr>
<td>Posted Note</td>
<td>229,603</td>
</tr>
<tr>
<td>Sticky Note</td>
<td>3,070</td>
</tr>
<tr>
<td>Other</td>
<td>7,449</td>
</tr>
<tr>
<td>Meeting</td>
<td>2,231,723</td>
</tr>
<tr>
<td>Appointment</td>
<td>795,223</td>
</tr>
<tr>
<td>Delivery Report</td>
<td>3,083</td>
</tr>
<tr>
<td>Read Report</td>
<td>215,282</td>
</tr>
<tr>
<td>Other Report</td>
<td>138,419</td>
</tr>
</tbody>
</table>
Why the stats blew out - 2013

• “Undeliverable mail” alert sent by the Exchange server - repeated identical messages over 2 days

• 260,785 in total - 55% of emails for that month
Why the stats blew out - 2014

- “Undeliverable mail” alert sent by the Exchange server - repeated identical messages over 3 days
- 448,587 in total - 65% of emails for that month
Why the stats blew out - 2015

• “Undeliverable mail” alert sent by the Exchange server - repeated identical messages over 3 days

• 149,116 in total - 28% of emails for that month
Why the stats blew out - 2015

- “Test” email sent by the web server over 13 days – each with an attachment, totalling 3.2Gb
- 94,209 in total - 18% of emails for that month
Are they unique or duplicates?

• If the emails are duplicates then we can dispose of them without further ado.

• How do we tell if they are duplicates?
  – What other systems may have captured them?
  – Are they accessible within those systems?
  – Are they being managed (retention assured, disposal applied) within those systems?
  – Can we have absolute confidence that they have been captured in record-keeping systems? How can you tell?
What systems might they be in?

- They may be accessible in the individual staff email client – but inaccessible to others and vulnerable to deletion.
- They may be held in the Exchange Server – unless deleted through the desktop client – but not accessible except to original sender/recipient and do not have retention criteria applied to ensure they are retained - duplication of what is in the Archive.
- They may be in an EDRMS or a printout on file.
What sort of junk items are held?

- System alerts (daily averages)
  - 70 emails from the automated library sorting machine
  - 50 library alerts to customers – items due, reservations
  - 100 emails for Helpdesk requests
  - 35 voicemail alerts
  - 10 Twitter, 35 LinkedIn, 10 Pinterest notifications, 35 Google Alerts (not business use)
  - 75 Exchange Server notifications
- Over 420 junk items of only short-term value per day – 153,000 per year – and that’s only some of them
Disposal based on ALGIM and GDA7

- Calendar items – delete [ALGIM 23.6.1]
- Delivery and Read reports [GDA 7-1.6]
- Leave requests/approvals [ALGIM 25.6.1]
- Travel bookings [ALGIM 12.6.1]
- Helpdesk requests [ALGIM 27.1.1]
- Out of office notifications [GDA 7-1.6]
- Etc. . . which disposes of millions of items
Identifying Emails for Disposal

• Based on
  – Email message type (Read report, Calendar item)
  – Sender – OneDaysale, TradeMe, Booking.com, Expedia, Grabaseat, Flybuys, etc.
  – System notification – Spydus, Confirm, MS Exchange alerts, scans on MFD (14,000 for 2015 alone with attachments), Voicemail, TRIM alerts, etc.
  – Subject line – Request for Service, Seek, Sale, Friday Funnies (10,278 items), Newsletter (151,972 items)
  – Message body text indicating junk
Junk emails identified by message body

- “Don’t miss out” – 114,634 items
- “Buy now” – 61,764 items
- “Register online” – 43,613 items
- “To unsubscribe” – 431,685 items
Once we have disposed of the junk... 

- What to do with the remaining millions of emails?
  - Leave them in the email archive? But how can they be accessed when required and protected from unauthorised disposal?
  - Migrate them into the EDRMS? A huge job given the difficulties in extracting them from the archive, and no easy way to attach them to the records classification
  - Store them in a separate area in the EDRMS? How do we arrange access given the risk of exposing private or confidential material?
  - If we restrict access to them because of the risk of breaching privacy or confidentiality, have we actually improved the situation?
Where to from here for the Email Archive?

• Once we have addressed the problem of capturing record emails into another system, what do we do then?
  1. Keep using the Email Archive and set up a process for ongoing management (capture to a record-keeping system, and disposal of unwanted material)?
  2. Discontinue using the Email Archive and ensure emails are captured as records via another method?
Considerations around discontinuing

• Is the Email Archive achieving what it was originally set up to do? (Does anyone even remember what that goal was?)
• Is it the most effective method of achieving that goal given widespread availability of EDRMS systems?
• Is there a better product which could be implemented which enables management more readily and effectively?
• Would that bypass the “archive” approach and capture emails directly into the EDRMS?
• Is that simply relocating the problem or could the emails be managed better inside an EDRMS rather than an “archive”? 
Deficiencies in our Email Archive tool

- Not designed for mass access to emails – search limited by system capability and impact on server
- Only two methods of exporting emails
  - Export to PST file which is then imported into Outlook
  - Restore emails to Outlook account
- Both cumbersome and designed only for small numbers of emails
- Not designed to enable RM – once emails are in the archive the assumption is that they will stay in.
- Searching inefficient due to the quantity of junk needing to be assessed for matching search criteria – the more stuff you need to search through the slower it is to return results
Considerations for continuing use

• Modify the journaling process to filter out calendar items, out of office notifications, etc. thus reducing junk captured

• Monitor capture of emails into the EDRMS and address areas where staff are not capturing their emails (through comparison of what is in Archive vs. EDRMS)

• Set up automatic capture of some emails into EDRMS e.g. all emails sent by CEO or Solicitor
Conclusion

• We have named the elephant
• We have a process for attempting to capture existing business records from the email archive
• We are planning to implement a disposal regime for junk emails in the Archive
• We will be attempting to filter out junk so the Archive does not capture it in the first place
• Hopefully this will disarm the ticking time-bomb
Thank you for listening!

• Contact [Chris.Gousmett@huttcity.govt.nz](mailto:Chris.Gousmett@huttcity.govt.nz)