



And then it went horribly wrong...

What happened when we tried to
manage our email archive

Chris Gousmett, Corporate Information Manager, HCC
ALGIM Conference, Wellington, 14-15 May 2018

The story so far...

- Presentation to ALGIM 2017 on the problems found in our email archive
 - Approximately 30 million emails in our email archive
 - Continuing growth at approx. 35,000 per month
 - 2Tb of disk space consumed
 - Continuing requirement for extra server space
 - Difficulties in accessing and retrieving emails
 - Unable to manage the emails as records

First attempt to improve management

- Identification of email “types” for ready deletion
 - System alerts
 - Out of office replies
 - Trademe and Flybuys ads.
 - Etc....
- A formal disposal request was approved
- A sample group of emails selected matching criteria
- The “delete” button was pressed.....

Not quite but close enough...



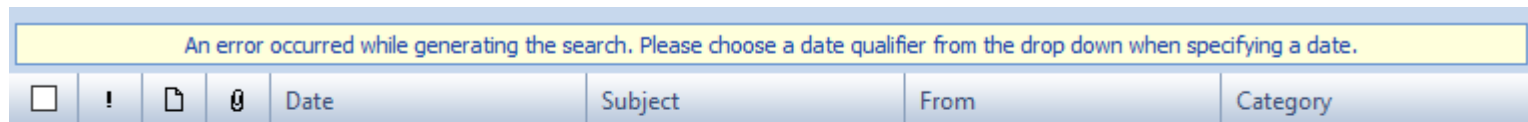
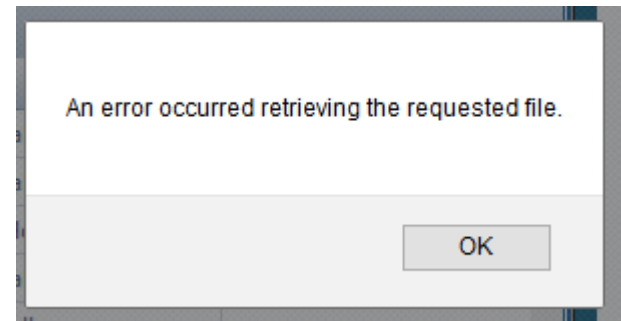
Result: chaos

- Everything appeared to work ok
- But then the help-desk calls began...

“I can’t search the email archive...”

“I get these error messages...”

“Why can’t I find my emails???”



No results to display

Enter your search criteria and click Find

Diagnosis begins

- The vendors were asked to identify the problem
- Outcome: the email system index was “in an inconsistent state”
- Vendor intervention was required
- Indexes were fixed and everything was back running as expected
- So then.....

The second attempt

- A selection of emails made
- The “delete” button was pressed

.....

This is what it felt like...



Result: more chaos

- The helpdesk calls started again...
- Vendor called in, same problem, “indexes in an inconsistent state”
- And so...

The outcome

- We were told to cease attempting to delete emails as the system was not able to process the delete request properly
- But all we had done was select emails and press the delete button on the toolbar!!!



Conclusion: Failure

- We were advised that the “delete” function didn’t really work and to stop using it
 - It removed the electronic document
 - It left the record of the document in the system with no way to distinguish between these and ones with an email still attached
 - The indexes were corrupted
- Which left us with the problem: how do we deal with 30 million emails plus an extra 35,000 per month?

First option: find a replacement archive

- Turn to Google and search “email archive solution”
- Result: lots of useful looking hits
- In depth analysis: every system had this nice friendly assurance – “You can be confident of finding your emails since **no email can ever be deleted from the archive.**”
- Not what we wanted to hear!

We needed an alternative approach

- Commercial “email archive” products are designed for the American market
- Focus is on discovery for legal action
- Works on the basis that all emails are potential evidence and need to be kept permanently
- Not a good fit for NZ

Other Options

- Modify the journaling process to filter out calendar items, out of office notifications, etc. thus reducing junk captured – not available with our product
- 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) – very labour intensive
- Capture the emails into the EDRMS in bulk

Capturing emails into the EDRMS in bulk

- Current approach

Exchange Server → Email Archive

This captures all email items into the Archive

- Approach currently being explored:

Exchange Server → ControlPoint → EDRMS

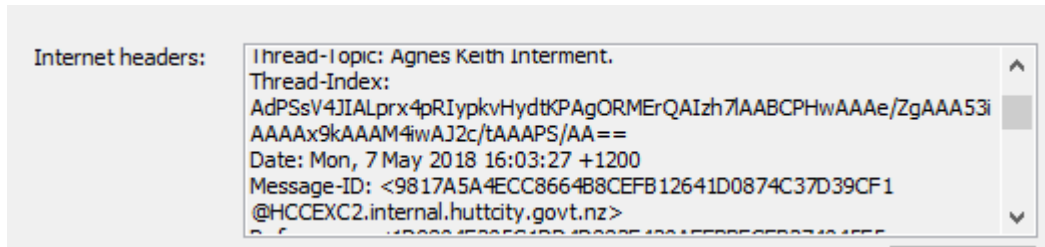
ControlPoint (from Micro Focus) has filter tools to allow elimination of junk so that only (mostly!) business records are captured.

Filtering

- Filters in ControlPoint would be based on the initial analysis of the Email archive contents to exclude e.g.
 - Calendar items (meetings and appointments) – 350,000 per year
 - Delivery/read reports – 45,000 per year
 - System alerts, promotional (TradeMe, Flybuys), social media (Facebook, Twitter)

Email ID issues

- Exchange adds an email header with unique ID to all emails



- Emails in the user's "sent" items do not have the ID as it has not passed through the server
- The EDRMS cannot identify it as unique and so rejects the email (if "unique identifier" is required) – this prevents capture of duplicates in the EDRMS
- "Sent" emails captured from Exchange Server have an ID

Business Emails automatically captured

But wait... there is always a catch!

- Not all business emails should be accessible by any and every staff member
- Not all emails will be business emails – private (personal) emails will be included
- Emails are set to be single-instance only – staff trying to save emails into the EDRMS will be prevented from doing so

Whereabouts in the EDRMS?

Given that there are privacy and confidentiality issues around capturing emails, the best option seems to be:

- Capture all emails into a folder in the EDRMS matching the user's mailbox.
- Users have access only to their own "mailbox" folder (EDRMS Administrators have full access) but in certain circumstances managers could access email folders for their staff
- Emails would need to be moved into the classification or marked for deletion

Where should the emails go?

- The preference of course is for the emails to go into the appropriate location in the business classification
- This requires Auto-Classification – available for our EDRMS but at a cost
- The alternative is to expect staff to move their emails into the appropriate classification – is this just moving the problem from one system to another?
- At least if everything is in the EDRMS we have the tools to manage it

Issues

- We *think* this might work
- We don't know of anybody else who has tried it
- Will users bother to move emails into the classification?
- Can we set the filters to be “just right” – capturing all the good stuff and excluding the junk – or will we miss important material, or capture too much junk (or both!)?
- Can we easily apply disposal criteria to emails not moved into the classification (say after 6 months)?

Issues

- Some staff have been proactive in filing their emails
– will they stop doing so?
- Global searches for legal discovery, official information requests, etc. will still need to be carried out by Records staff
- Configuration will be complex and will depend on the filter criteria and ability to direct emails to the correct user's folder

Conclusion

- Capturing emails into the EDRMS seems the only viable option
- We will be attempting to filter out junk so the EDRMS does not capture it in the first place
- We would implement a disposal regime for emails in the Archive
- It's still a risky experiment but it *might* work
- Hopefully this will disarm the ticking time-bomb (this time!)

Thank you for listening!



Contact Chris.Gousmett@huttcity.govt.nz