Victorian Government Email Machine Assisted Appraisal

Proof of concept
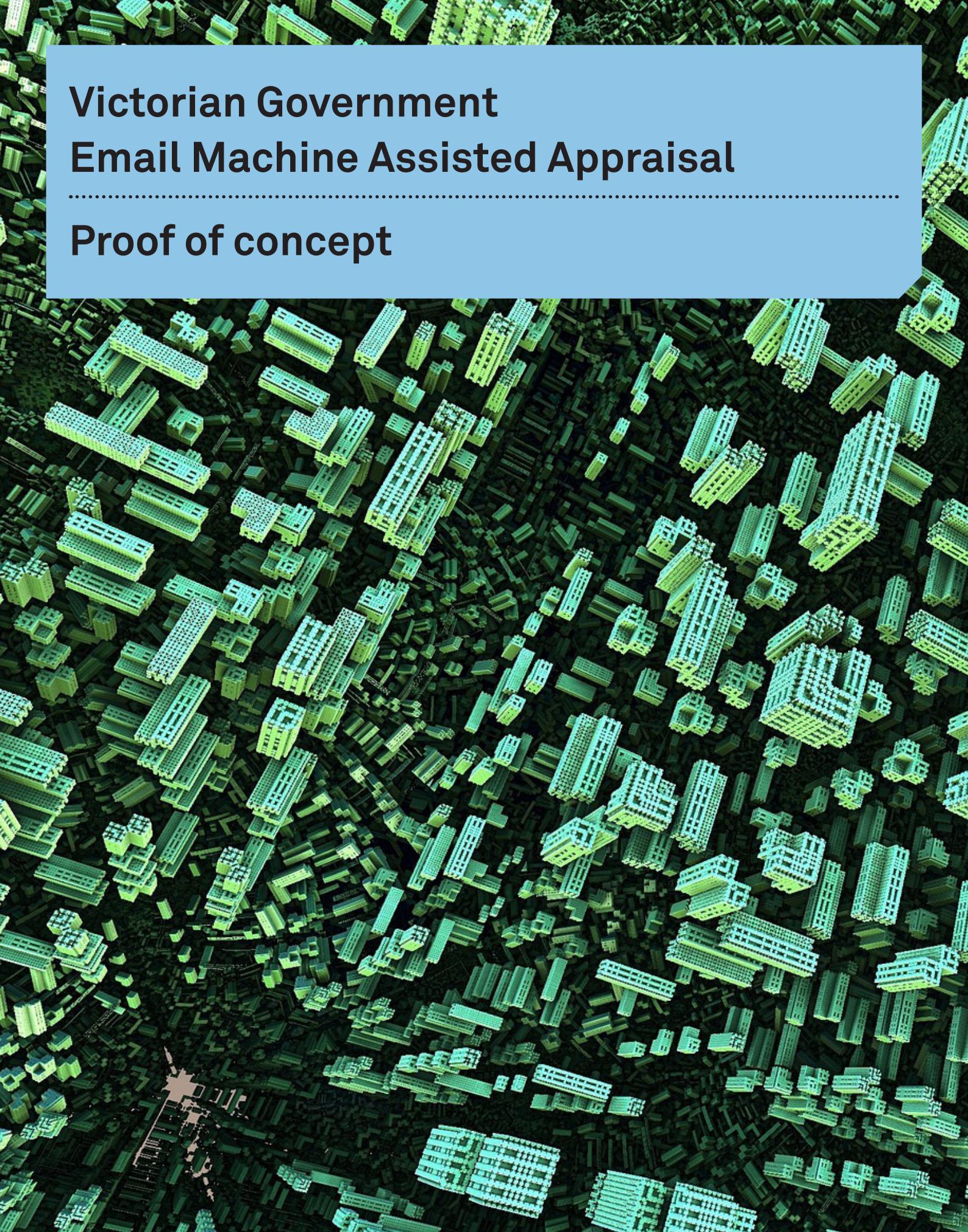
Emails are a vital part of doing business. In years past, important correspondence was paper-based and easily preserved. Now, key correspondence takes place over email. Emails enable exchange of ideas, enactment of decisions and support collaboration between an increasingly dispersed workforce. In government, emails also provide evidence essential for accountability and need to be preserved as public records into the future.

The problem

Over twenty years of routine backup has resulted in an unwieldly backlog of Victorian Government emails amounting to 67,000 tapes and 28 petabytes of content[1]. Access and retrieval of emails for the purpose of analysis and evidence of decisions can be difficult, expensive and time consuming. This compromises the Government's reputation for transparency and accountability.

Incomplete and ineffective retrieval of emails causes substantial and costly delays.

For instance, a 2012 VCAT ruling highlighted that email retrieval for one particular FOI request could amount to up to $970,000.

Every day the emails remain unmanaged, the costs and risks associated with management, storage and retrieval increase.

The proof of concept

We have been working with the Victorian Government technology provider, CenITex, on a project to make the Lotus Notes email backlog more available, valued and better managed. The Lotus Notes Proof of Concept (PoC) is the first step.

The PoC involved exploring the use of an eDiscovery tool to review and facilitate disposal of large volumes of emails, including:

- An initial assessment to quantify and qualify a sample email data set.
- Identifying duplicates within the data set.
- Identifying low value versus high value records within the data set.
- Assigning contextual information to the de-duplicated set.
- A manual review of results to determine level of accuracy.

The cost of storing emails increases over time while ease of access decreases.

Of a sample of 4.6 million emails, we found:

43% duplication of sample emails

1,678,633 emails

2,960,067 emails

- Originals
- Duplicates
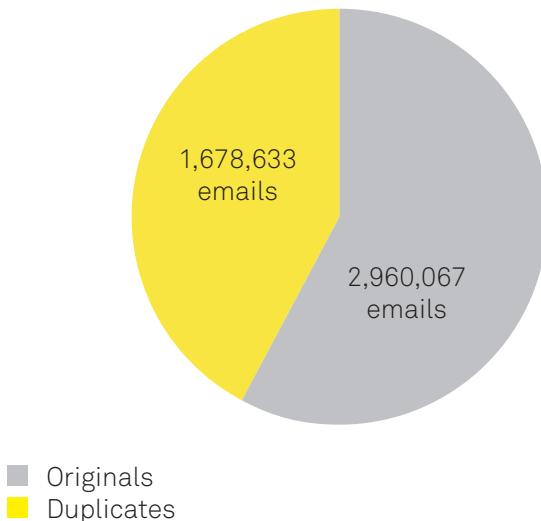
We found 7% were low value emails

2,765,339 emails

194,728 emails

- Originals
- Duplicates
- Material/valuable
- Non-material/not valuable
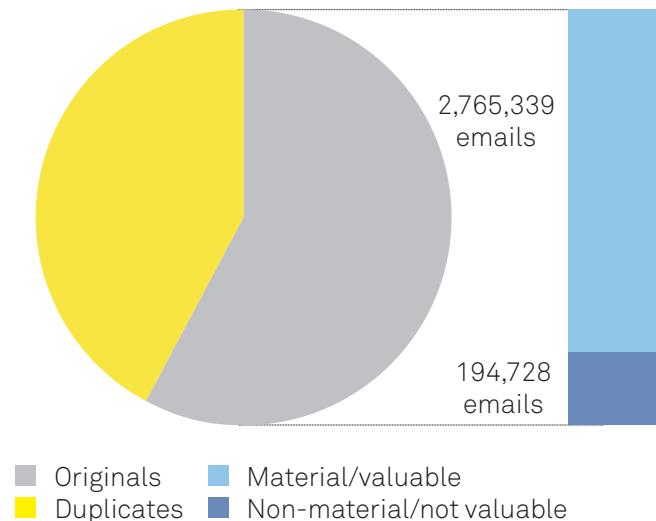
How we did it

Our goal was to reduce the volume of the email backlog in an authorised way; which in the Victorian Government means in line with Retention and Disposal Authorities (RDAs).

The tool was used to identify duplicate emails from within the sample.

To identify low value emails among the remaining sample we reviewed a list of email domains to identify those that would reasonably result in irrelevant, non-business related emails. The top results, which included common subscription emails and Google alerts, were selected and saved as filters. The use of Fwd: in the subject line was also used as a filter.

Next we tried a second approach on the sample, searching the remaining emails for key search terms.

Using a third approach we were able to apply additional contextual information to the emails, which would allow them to be grouped by areas of responsibility within the organisation. This allows us to assess and prioritise the emails to be kept long term.

The findings

The eDiscovery tool was successful in allowing us to identify emails eligible for disposal, as well as assessing and prioritising remaining emails with between 98% and 100% accuracy.

Up to 50% of the sample was identified for potential disposal.

The tool allowed us to apply additional metadata to every email in the set, enabling easier identification of emails at a high level, facilitating future decision making around retention.

In summary

An eDiscovery tool may be used to assist agencies to reduce their email backlogs and unlock greater value from their email assets, though a larger sample of manual testing is recommended prior to implementing disposal.

An eDiscovery tool may be beyond the means of smaller agencies who nonetheless struggle with similar email backlog issues. An investigation into email back-up for smaller agencies and potential testing of free, open source solutions is recommended.

[1] Landell Consulting, 2013. CenITex Lotus Notes email management Processes. Melbourne: Landell Consulting