Preserving Email

Digital Preservation Topical Note 7

DigitalPreservation**Coalition**

Why is preserving email important?

Since the first email was sent in 1965, it has grown to become one of the most commonly used forms of communication for both personal and business interactions. As much of modern correspondence takes place via email, it can provide important evidence of key transactions and details of decision-making. Therefore, preserving access to emails can be essential to ensuring the accountability of those in power. Access to emails has already proved a crucial part of uncovering illicit business practices, e.g. the Enron scandal in the USA. It is also an important historical and social record, with researchers already demonstrating they can construct pictures of social and business communities through their email interactions.

What risks do emails face?

Email is a format that is characterised by ubiquity but also ephemerality, both of which can cause problems for long-term preservation. These risks include the following:

- Due to the volume and nature of email, it is often treated with benign neglect (see note on Personal Digital Archiving) or gets deleted to clear space without proper consideration of the importance of the content.
- Most organisations lack clear policy and guidance for email users on how to manage and preserve email, meaning that everyone takes a different approach.
- Email correspondence often contains proprietary, sensitive or private information so providing access can raise issues with regards to Freedom of Information and Data Protection.
- Emails can be incredibly complex, containing references to external content which is required to understand the message, as well as allowing almost any file type to be included as an attachment.
- An email from a longer chain may be rendered meaningless if the chain of messages and surrounding context is not retained.
- Email is particularly susceptible to accidental or malicious deletion, e.g. it is often a top target for hackers.
- While the format for the transfer of email between systems is standardised, the formats used by the software used to access and manage email is not. This means email formats can be dissimilar and different access portals may store emails in disparate locations.

www.digitalbevaring.dk

Who is responsible for email?

More than with any other type of digital format, the responsibility for the active management of email sits with individual users. We all manage our own inbox, choose which emails to retain, and assign folder structures. We also often feel freer to delete an email than we might other types of digital document. It is, therefore, important that we recognise this responsibility and take a methodical and considered approach to the management of our emails.

However, not all of the responsibility for access to and management of email lies with the user. An organization's IT Department will have responsibility for installing and maintaining the email client used, as well as storage and user accounts, and potentially an exchange server used to transfer email. Records management and/or archives staff may also have responsibilities for creating email retention policy and guidance (and it is worthwhile checking to see if these exist for your organisation). Archive staff will also be responsible for the long-term preservation of emails when they are no longer in active use.

www.digitalbevaring.dk

Key Term: **XML**

XML (eXtensible Mark-up Language) is a mark-up language used to give a machine-readable structure to information. Like punctuation helps us to understand the structure and meaning of written text, XML tags allows computers to 'read' and understand electronic documents. For example, HTML is a sub-set of XML and shows browsers how a webpage should be structured and displayed.

How can email be preserved?

Printing emails might seem like an easy or obvious option, but this can lead to the loss of a large amount of information. Attachments and context in relation to other messages (such as threads) may be lost as well as the contents of email headers which contain rich information. Email can be preserved as a single message, as groups, or as the content of a whole user account. This is most commonly migrated to a system-neutral format using one of a variety of specialist tools that are available. These neutral formats are usually *XML-based*. Information about an email's context/ relationship with other emails will also be captured.

What steps can be taken to facilitate email preservation?

There are a few simple steps that can be taken to help facilitate email preservation. These include:

- Avoiding 'preservation hostile' configurations of email servers and clients. This includes turning off any auto-deletion functions and making sure there is enough storage for the volume of email received.
- Adhering to any policy or guidance relating to email use within your organisation.
- Keeping all email in one account, rather than forwarding between different work and personal accounts.
- Plan for how you will manage your emails, including setting up folders for filing important emails and doing this on a regular basis. Some systems may also allow you to 'tag' emails which can help generate metadata at a later date (see note on Metadata).

For more information on Digital Preservation visit the DPC Website: https://www.dpconline.org