

## Email appraisal: developing strategies to tackle a complex problem

“the problem won’t take care of itself, and the time to act is now.”

This quote comes from the Council on Library and Information Resources which published a report on the preservation of email in August 2018 came at an excellent time for me, having just begun in my role in digital preservation at the University of Warwick. The report describes itself “as a call to arms to push the community to take action on the defined working agenda by describing why email archiving is a compelling and sound investment for modern historical records and research.”

It goes on further to say: “The most important work to be done in advancing email preservation lies simply in nurturing and fostering archives and libraries that are leading the work or wish to become more fully engaged. In theory, every archive that is collecting or preserving contemporary record material is collecting and preserving email, making clear the need for **increased knowledge, information sharing, and collaboration.**”

I am trying to answer that call and would be delighted to hear from anyone here who is interested in doing more or sharing workflows, practice or just wants to chat about what they could do.

So - is there a problem? I don’t really like the word problem as it implies there is a thing to be “solved” here. The ultimate goal of preserving emails - and indeed with any archival preservation activity, digital or otherwise - is not something which is done or solved but rather an ongoing iterative process which is developed and refined over time but never reaches a conclusion. It is not set in stone but new theory and practice emerge over time and we refine and develop our own practices accordingly. But the fact is that there are some specific problem areas which face the archivist when undertaking their work and I would like to explore some of these issues - with particular reference to email appraisal - in this paper, with the aim of encouraging others to share their own work, plan for what they might do and help shape community best practice in this area.

Emails have been an integral part of the way we communicate for over twenty years now. I sent my first email in 1995 and dread to think how many I’ve sent - or worse still received – since then. They have helped keep me in touch with friends in faraway places as well as help with planning parties, holidays and family events. I also spend a huge percentage of my work day reading, answering, filing and deleting emails. Emails have become almost synonymous with work for many people, which I think feeds into some of the problems we face when trying to deal with them in an archival context.

And the funny thing is that as we sigh over the inbox with 250 unread messages we are still sometimes being told that emails are obsolescent.

To be fair to this journalist (<https://www.independent.co.uk/life-style/gadgets-and-tech/features/are-we-facing-the-death-of-email-6273170.html>) he rightly noted that the use of email in the workplace in 2011 was very significant and increasing and other communication methods remain in some ways tend to stay more in the sphere of personal communications such as - Slack, WhatsApp, Yammer etc etc. These communication tools have their uses - especially as collaborative platforms in an age of home and flexible working but most people generally recognise on some level that they do not have the same kind of evidential weight as emails - perhaps because there is something more formal about the email with its addresses and subject lines and so on. Perhaps even (depending on the client) the ability to file messages. In fact these qualities plus also - in a very big way - the fact that unlike the instant messaging or collaborative tools I have just mentioned email is not a proprietary format should make tasks such as appraisal, curation and preservation easier. There are international standards set by the Internet Engineering Task Force

(IETF) which email clients adhere to - otherwise we would not be able to communicate across platforms and service providers. This is actually helpful in that allows us to develop strategies based on already agreed standards in order to capture meaning, context and metadata associated with email format.

Do we think of emails as a little bit ephemeral? It definitely encourages informality so I think it is wrong to think of emails as a “replacement” for traditional letter correspondence (although in some fairly obvious functional ways they are). But more than that - a place where work happens (ironically I think this is the strapline for Slack, the aforementioned organisational communication tool, which some might have thought would replace emails but... well it hasn't, probably because it doesn't offer all the things email does. It's a message service but may be used in different ways - the message may be to yourself - I know I do this quite a lot - or the message might only make sense in relation to the attachment. The message might be contained within a calendar invite.

“Email is not one thing, but a complicated interaction of technical subsystems for composition, transport, viewing, and storage.” (CLIR)

Before we embark on any approach to appraising email content we need to understand what email is and how it works. The full header information, part of the delivery method, offers far more than the envelope and postal mark on it in terms of contextual information. The current standards which exist mandate that certain data fields are available to ensure successful delivery but these do not apply to the storage of email which are not governed by any standards - in other words each proprietary or other other mail clients - can and does do things slightly differently. This has implications for us when we are making appraisal decisions as one “version” of an email sent by one person and stored in their outgoing mail folder will not necessarily be the same as the received version another person has in their incoming mail folder. There are also no mandated standards for how mail clients deal with attachments. Emails are fluid form -they can exist at any time in several mailboxes, in different proprietary formats, as a sent copy a written copy, a forwarded copy etc. With all this to consider which of these might you choose to capture?

So what is happening on the ground at the University of Warwick where I work? Our University is not an ancient institution - we recently celebrated our 50th anniversary but with this it came an interest and ownership in the idea of heritage - it's certainly become easier to sell this as a concept as the institution adjusts to the idea of being “historic”. Alongside this changes in information legislation in the United Kingdom meant information governance was pushed higher up the agenda and brought with it a Records Manager - my colleague Mark - without whom my job would be impossible. The key to any successful institutional digital preservation programme is resource and investment - we finally had resources to devote to management and preservation of digital records and also a recognition of the value and importance of the work.

We were then promised two email inboxes belonging to previous senior officials of the University. This would complement very nicely the existing collections which we have representing their predecessors. I had hoped that this paper was going to be me sharing the results of both this deposit and how we tackled the appraisal of it. Alas - as is often the case - the inbox has fallen foul of what I see as the primary issue in digital preservation (before appraisal even comes into it) which is actually capturing the material in the first place. We have made great strides in capturing particular classes of corporate record - notably our committee minutes and papers we have done less well in other areas.

In the case of emails I think there are several things causing this issue:

- People don't recognise them of being of either legal or historical value

- Our own lack of resource to either take digital or to have anyone with oversight of the University's Archives (no Records Manager or University archivist).
- Reluctance to share things where there is a perception (and indeed reality) of a mixing the private and the public.

So what are we going to do about this? The Council on Library and Information Resources report says "*In theory*, email should be controlled by a well-defined records management system and appraisal will be guided by established records disposition schedules" (CLIR). But also recognises that this almost never happens. The University of Warwick has an established records disposition schedule and it usually prompts the comment "but it doesn't mention email", as if emails were a class of records in their own right. We are starting to look at how we might be able to use Sharepoint and Office 365 to guide retention and capture of records - including emails - but this is a very long way away from becoming a reality not least because of the excessively fragmented and devolved environment.

We need an explicit statement of our interest in emails as part of records environment and a recognition of them as being of legal and historical interest. Although the archivist in me dislikes this identification as the record in terms of its format it might counter the confusion and help with raising the profile of the activity. We need to frame email capture as part of a continuum already established with the current holdings which include the correspondence of previous officials of the University. We need to establish both legal authority (which we have) and institutional mandate (slightly less sure about this) to collect and appraise emails.

At the US National Archives (and other institutions) they have adopted the Capstone approach. This is where the key/most important office holders are identified and all their emails are captured. I have often felt it is the only one which is going to capture that sense of what was going on in an organisation including context and background. However in the UK this approach is very difficult to square with the legislative framework and I do not think my organisation is ready for this (either yet or ever)

There are a number of possible approaches:

- Rolling capture - an ongoing appraisal process which has the advantage of capturing records soon after creation but will lead to duplication and can't actually imagine this happening
- Let the creator do the selection and appraisal
- Keep only the sent box (don't like this especially but it has a certain pragmatic appeal).
- Keep everything (I don't like this approach either at least not for institutional records - a different matter entirely for literary type archives). The problem with keeping everything that this is almost dereliction of duty - it isn't appraisal at all. And there are the obvious legal implications (see above) and then also environmental impact, resourcing etc

We will probably be taking a mixed approach. At present - because we will take what we can get - we will guide the creator (or in this case the secretary to make the selection before depositing using key words relevant to the custodian's knowledge of the content. Even this approach - requiring time to prioritise what for the secretary is a very low priority task - is likely to prove difficult. We have to live with the fact that there is no neutral appraisal and any selection criteria are nuanced and biased.

Does this mean we should not conduct appraisal? Absolutely not as it is part of our role to manage both the records themselves and the resources we need for the task. The key thing is to document the processes that we use and the rationale behind any decision making.

So far so good but if we want to take this forward we want to find more sophisticated tools which can help with the job...

EAdd was developed by the University of Stanford as an open source tool to aid the appraisal and discovery of emails. I love ePADD - I have blogged about my failure to use it here but I am pleased to report that since then I have come some way and now have it up and running successfully.

It is a great tool which was developed for and by archivists, meaning it is easy to use with a straightforward user guide, a supportive online user group forum and a responsive admin support. It only supports email inboxes saved as the MBOX format which is quite widely used, although inevitably not one used by Microsoft Outlook, which our University uses. However:

1. It is possible to convert to MBOX using a number of tools eg Emailchemy or Aid4Mail are two such - none are free but there are trial versions to play with
2. In any case this was a theoretical problem as I hadn't actually got an inbox to play with so for the purposes of actually testing out a workflow I used my own personal inbox which happens to be a gmail account (which uses MBOX). So no conversion required.

There are a few things you should know about me and one of them is that I am one of those dreadful people who never deleted anything from their inbox. I am the records manager's nightmare although I was also able to produce a useful sample of emails of a large sample size and probably analogous to what I might expect to find at the University. After all I know from anecdotal evidence that the general trend at the University is never to delete anything ever - very poor practice especially from an information governance point of view but possibly better than the other extreme from an archives point of view of deleting everything.

I'm going to encourage you all to go away and have a play with ePADD if you haven't already because it offers much for the archivist particularly with tackling vast amounts of data.

First of all there is a generic report which highlights errors including:

- Undated message
- Unnamed attachments
- Duplicate messages
- Duplicate attachments
- Unsupported encoding

These error categories can be navigated from the main user interface

It gives each message a unique ID which is really helpful when searching, annotating and deleting. You can also label messages with different review and restriction features.

You can get some great data visualisations such as this one which is "emails my mum sent me over the last 10 years" - and you know what that spike here is? That's when we bought her an iPad and she started sending emails more...

What else is required? A need for robust and clear policies on emails which will outline where responsibilities lie and what procedures will be followed. What I want to do is build up both our own confidence and our own profile so that saying things like "we want your email" doesn't sound strange, unexpected/unwelcome etc. This will mean building bridges with a number of different people including our IT services, with whom it can be difficult to get traction with.

The Council for Libraries and Information Resources report identifies the perception of email capture as a difficult problem standing in the way of progress and suggests the development of ad hoc workflows and duplication of effort between institutions has slowed the momentum of the prioritization of email capture and appraisal and rather negatively suggest a vicious cycle of increasingly complex problems which remain unresolved.

I want to end on a more positive note. I recognise that progress in any aspect of dealing with digital records is very slow but there are things we can do. We can get out and about talking to people about what we are trying to achieve; we can spend time planning and mapping out our procedures, writing our policies and documenting our decisions and we can and should share what we do. We can share workflows and best practice via a wide variety of platforms - different ones will suit different topics, audiences and people but the more we shout about what we are doing and the more we can work towards community best practice the better for all of us.

Rachel MacGregor

rachel.macgregor@warwick.ac.uk

#### Bibliography:

Pennock, M. 2006. "Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages."

<http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>

Bearman, D. 1994. "Electronic Evidence: Strategies for Managing Records in Contemporary Organizations" Pittsburgh: Archives & Museum Informatics.

[http://www.archimuse.com/publishing/electronic\\_evidence.html](http://www.archimuse.com/publishing/electronic_evidence.html)

Council on Library and Information Resources (CLIR). 2018. "The Future of Email Archives: A Report from the Task Force on Technical Approaches for Email Archives"

<https://clir.wordpress.clir.org/wp-content/uploads/sites/6/2018/08/CLIR-pub175.pdf>

ePADD User Guide 2019

[https://docs.google.com/document/d/1CVIpWK5FNs5KWVHgvtWTa7u0tZjUrFrBHq6\\_6ZJVfEA/edit](https://docs.google.com/document/d/1CVIpWK5FNs5KWVHgvtWTa7u0tZjUrFrBHq6_6ZJVfEA/edit)  
(accessed 18/7/2019)

Lappin, J. 2019. "Predictions for the application of machine learning to the management of email".

Blog. (<https://thinkingrecords.co.uk/2019/02/26/predictions-for-the-application-of-machine-learning-to-the-management-of-email/> accessed 17/7/19)

The National Archives (TNA). 2013. "What is appraisal?"

<http://www.nationalarchives.gov.uk/documents/information-management/what-is-appraisal.pdf>