

Case Study 2: Email Management and Preservation

UK Context

Dr. Anthea Seles

6 July 2017

Introduction

- The question of where the challenge lies for email. Is it really about preservation?
- Government realities around information management and its impact on email management
- Can we turn what we perceive as the challenge for email into an advantage?

The question of where the challenge lies for email. Is it really about preservation?

- The challenge isn't preservation, the challenge is capture and presentation
- Capture
 - Identifying what is of value
 - Identifying what is sensitive
 - Dealing with pst files
 - Remapping existing relationships between unstructured records
- Presentation
 - Presenting email
 - Meaningfully presenting email for researchers to use

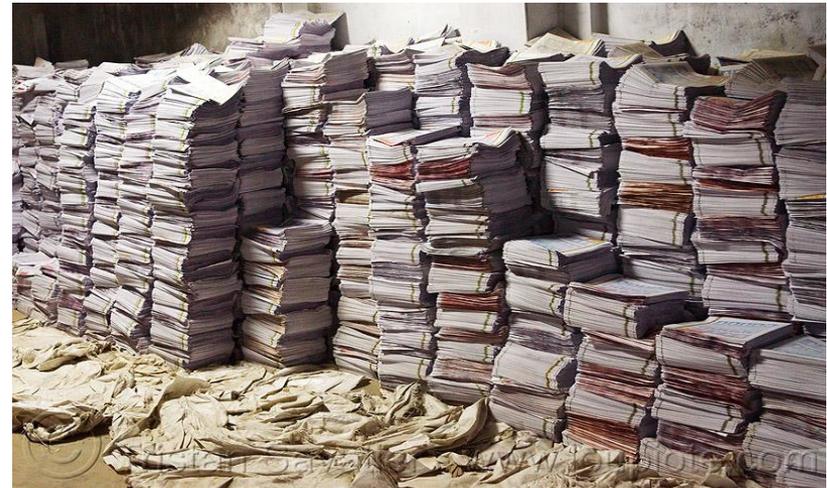
Digital Realities in Government

“In some government departments,

for every TB of information held in an EDRMS, there are approximately 10TB of data in shared drives.

Even when an EDRMS exists, shared drives often exist in parallel. Shared drives often have no type of record-keeping controls in place to limit or monitor filing ”.

‘Digital Landscape in Government 2014-15 Report’
by The National Archives



Digital Records Realities in UK Government

- The challenge with email, as with unstructured records collections in government is identifying what is of value and then what is sensitive.
- The challenge exists with structured collections (e.g. information management systems) as well Two email realities in government:
 - Just drag and dropped large volumes of email in information management systems. The result is that 60-70% of the content of information management systems were email
 - Dealing with pst files
- Where there has been conscientious use fo information management systems pointers included in emails cause challenges as well
 - Time-space question (i.e. document it is pointing to has been over written)
 - 404 Not Found errors

Presentation- Discoverability and Accessibility

- Our major challenge now is how best to present emails to our users
 - When taking into account sensitivity review
 - When taking into how researchers want to consume and re-use for their work
- ePadd by Stanford University offers us a way forward and an interesting approach to discoverability over accessibility



Can we turn what we perceive as the challenge for email into an advantage?

- It's where all the meaningful conversation happen and where we can identify major events or topics
 - Ex. Carcanet Press- University of Manchester
- Can we potentially use email to extract entities and subjects that can be used to train a machine to identify meaningful records in unstructured records stores?
- Is there therefore not more merit in Capstone where we consider the most topical conversations happen?



Conclusion

- The issue isn't preservation, it's capture and presentation
- Email is problematic but can we flip the problem into an advantage
 - Understanding unstructured collections of records
 - Visualisation of email to help with sensitivity review
- Managing discoverability over accessibility, there is also a broader question for other unstructured records