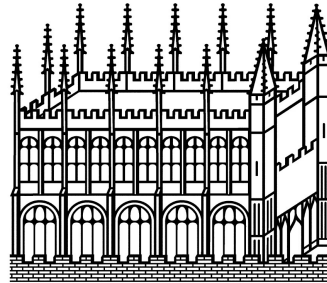


Receiving and managing email archives @ the Bodleian Libraries



Bodleian Libraries
UNIVERSITY OF OXFORD

Susan Thomas,
Digital Archivist/Project Manager
Bodleian Electronic Archives & Manuscripts (BEAM)



<http://futurearchives.blogspot.com>
<http://www.bodleian.ox.ac.uk/beam>

Operational context

- A 'collecting' archives unit
- Individuals & organisations
- 'Dead' collections
- 'Live' collections
- 'Hybrid' collections



Personal mailboxes – all shapes and sizes



<http://www.flickr.com/photos/joanet/5094833752/>

#1

Our first email archive

- Email on behalf of MEP/peer (staffers)
- EP Had an 'archive' system in place
- Low on structure
- 200,000+ emails
- MS Exchange/Outlook environment



<http://www.flickr.com/photos/xaf/2289377358/>

IT rules to manage email use

2. Rules governing mailbox use

To ensure the proper performance of the email service, Parliament has, since 10 December 2004, taken the following measures:

- 1) In the **Inbox** and **Sent Items** folders (including their subfolders):
 - automatic deletion of all messages older than **60** days.
- 2) In the **Spam High** folder, if you have one:
 - automatic deletion of all messages older than **7** days.
- 3) In the **Deleted Items** folder:
 - automatic deletion of all messages older than **7** days.

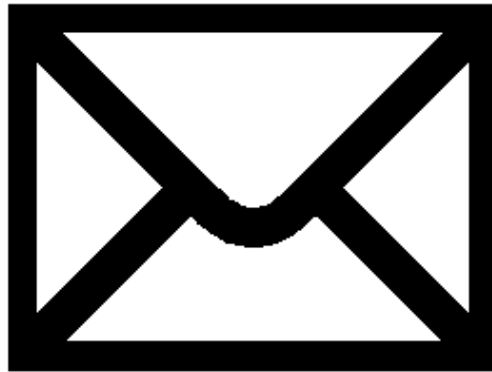
As of Friday, 25 February 2005 the following additional measure will apply:

- 4) **In all other folders** (including subfolders) of your mailbox:
 - automatic deletion of all messages older than **60 days**.

If you want to keep a message in your mailbox for longer than 60 days, you will need to archive it manually, using the function **'Mark to be EAS archived'** in the **'Actions'** menu of Outlook.

If you do not do this, your message will be automatically deleted after 60 days, whatever the mailbox you have filed it in.

Records survey (email-specific)



- Agree selection
- See system in action
- Obtain supporting docs
- Meet with EP IT re. practicalities
- Needed written authorisation from the MEP

Our first email extraction & migration

- EP IT export email from server side
- .pst file split over 5 DVDs (20GB)
- Delivery by secure courier
- Validate contents and acknowledge receipt
- Recreate single .pst for bit-level preservation
- MBOX migration with Thunderbird

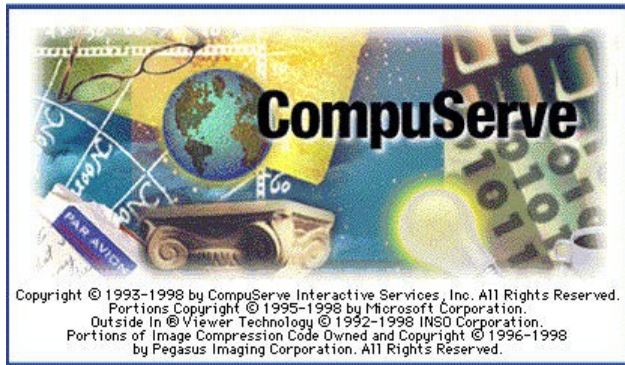


<http://www.flickr.com/photos/jonahsniper/2800041431/in/photostream/>

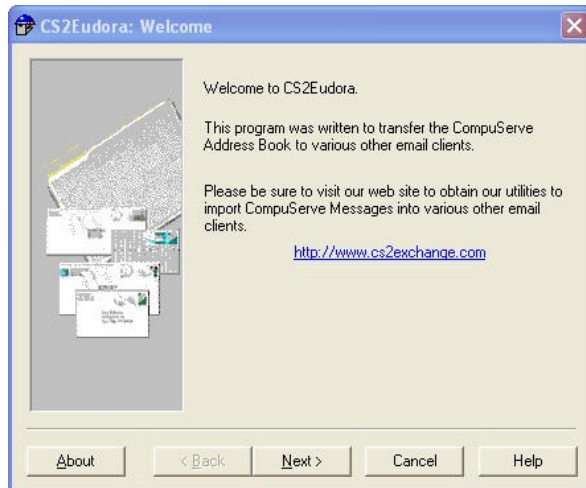
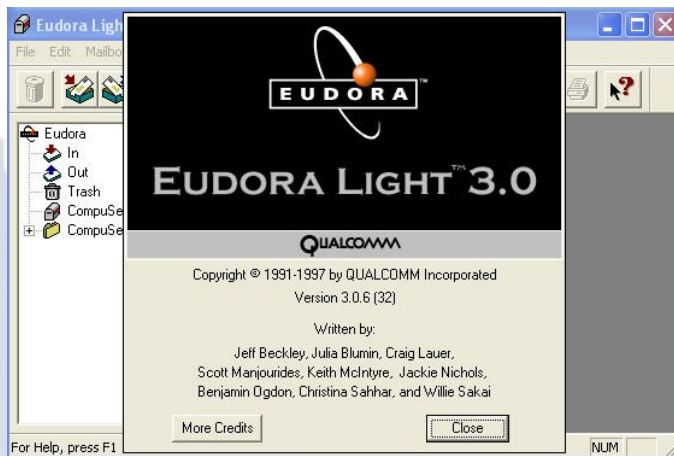


#2

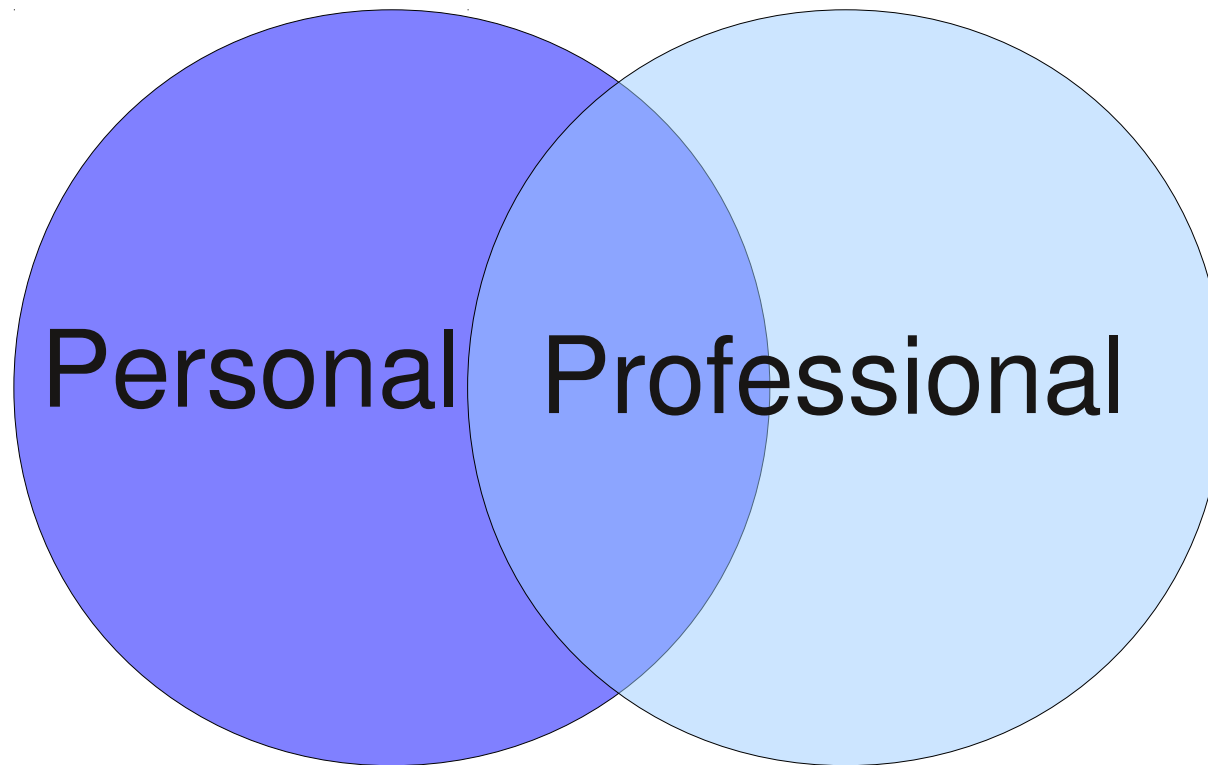
CompuServe 4.0 a/c



- Very old, but new to me
- Research
 - Where is the email?
 - What format?
 - How to migrate?
- We were lucky!
 - Old software still available.



Appraisal & selection



- Use FTK for review (tagging)
- FTK metadata export to facilitate family review

#3



Email from the cloud

- Book project – a 'dead' collection
- Simpler for us!
- Download via POP or IMAP
- Use a/c credentials
- Labels? (Download will replace these with folders)



<http://www.flickr.com/photos/9190330@N06/5810163712/>

Organisational mailboxes



<http://www.flickr.com/photos/harrychen/2736885773/>

#4

Incremental email capture

- Ongoing relationship with small press.
- Email arrangement facilitates snapshots - folder per book project.
- First transfer poor. Important metadata lost.
- Second transfer mediated – logical images of directories inc the relevant email.

#5

Organisational email - big scale



- Long-time user of Lotus Notes (.nsf format)
- Poor client-side export
- IMAP access theoretically possible. Practically not.
- Require assistance from organisation's systems dept. to acquire at server side (good will – this not their job!)
- Tools available to migrate to pst (require installation of Lotus Notes). Can migrate from there to mbox.

Some final thoughts

- Email capture – lots of variables!
- Need email characterisation tools
 - How much stuff is in the mailbox?
 - How is it organised?
 - How many attachments? Formats?
 - What's it all about?
 - Did we migrate it well?
- Aqua mashup - tools for mbox

Some more final thoughts

- So far, internal users only.
- Researchers will want
 - Visualisations
 - Client-like interfaces
 - Faceting/sorting using email metadata
 - Full text search
- Most email too recent for release into researcher domain
- In the meantime, other problems to solve...

FIN

Susan Thomas
Digital Archivist/Project Manager
Bodleian Electronic Archives & Manuscripts (BEAM)

susan.thomas@bodleian.ox.ac.uk

<http://www.bodleian.ox.ac.uk/beam>
<http://futurearchives.blogspot.com>