E-mail Archiving
All Things to All People

Garry Booth

Graeme Fowler

Carys Thomas

Where are we from...?

...and how on earth do you pronounce it?

Loughborough and the University

- Industrial & market town in the heart of the East Midlands, pop. approx 60000

- University established 1966, colleges 1909

- Single campus, 433 acres

- Approx 15000 under & postgrad students

- Approx 5000 faculty, staff, postgrad researchers

- Research intensive

- International renown for sporting achievements

- 2006 Times Higher "Best Student Experience" award, 2007 National Student Survey top 5 for student satisfaction

- Which country are you from?

- Does your institution use any form of email archiving?

- Does your institution have mailbox quotas?

What does "email archiving" mean to you?

- Cultural or historical research

- Take email off central servers onto users' hard disks

- Removal of "old" material off fast storage media

- Selection of emails to be preserved in accordance with retention schedule

- Keep a copy of all email for legal compliance - tamper proof (applies more to US than UK)

- ...something else?

Our Presentation Today

- Background

- JISC Funded Project (2003)

- Surveys of UK Universities

- Some Issues in Implementing Email Archiving

- Design of Loughborough solution

  - User consultation

  - Risk Analysis

  - Technical considerations

- Loughborough Implementation

Background - 2003

- Much University business conducted by email

- Individuals decided which emails to dispose of and which to keep

- The University backed up email for disaster recovery but when staff left, it was deleted

- UK Data Protection & Freedom of Information legislation affecting the management of email

JISC Project Bid - 2003

- Study of the Records Lifecycle - Specialist Electronic Studies examining Institutional Email

- Examine current working practices

- Develop policies for the retention and disposal of email

- Evaluate technical options for archiving email

- Inform University and UK HE sector of the findings

Main Tasks

- Carry out interviews with staff

- Conduct a survey of other institutions

- Draft an email retention and disposal policy

- Research technological solutions

Findings – Staff Interviews

- Email is an essential tool of communication
- Decisions and responsibilities agreed by email
- Job responsibilities affect email management
- Staff often kept paper copies of important emails
- Staff had different levels of skills and expertise
- Wide variety of practice
- Staff wanted control of any archive

Retention and Disposal Policy

- Template produced by the project for use by institutions
- Emails are owned by the institution
- Individual staff responsibility to decide what emails to archive – capture these as records and retain in accordance with retention schedule
- Institution responsibility to provide the technology to do it plus guidance and training

Project Deliverables

- http://www.lboro.ac.uk/computing/irm/index.html

- Resources include:
  - Template – institutional email disposal policy
  - Case studies – how users manage their email
  - Final project report

- Thanks to Michael Norris, Project Officer

Survey of UK Universities, 2003

- 21 institutions responded

- Only general policies were evident

- Respondents were backing up their email rather than archiving it

- No institution was identified which had a well defined email archiving policy

Survey of UK Universities, October 2004

- 26 Universities and Colleges
- 22 not archiving email
- 4 in progress
    - Two taking a copy of all email
    - Other two – focus was management of mailbox sizes

Survey of UK Universities, March 2006

- 28 Universities and Colleges
- 1 has implemented
- 7 in progress
- 20 not archiving email
- Movement towards installing such systems
- Focus is technical management of storage, not records management

Issues - Whose email is it?

- Does it belong to the institution or the individual?
- There are drivers for institutional ownership of email records
- But users see email as their own personal property
- There are issues of academic freedom and intellectual property rights
- When staff leave what happens to their email?

Issues - different styles

- Some people like to organise things

- Others don't …

Issues - Why bother?

- The business case is not easily made

- No high profile legal cases in our sector (yet)

- Doesn't appear to move anything forward

- No staff productivity gains (on the contrary)

- Staff indifference (at best) …

- … or resistance

- Zzzzzzzz

Issues - Management Buy-In

- **Needs senior management engagement for:**
  - Formal adoption of policies
  - Championing
  - Monitoring, enforcement
  - Funding
- **But it's not high on their list of priorities**
- **Is it clear whose responsibility it is?**
- **At Loughborough – agreed to promote guidelines, but not to enforce a policy**

Issues – "There is a bigger picture"

- Electronic document management systems encompass email as well as other electronic documents and scanned images of paper documents

- Undoubtedly "better" (more organised, more elegant)

- But more expense, complexity etc – the case is even harder to make

- Delays considerations for email

User Consultation - summary

- Staff want control over their own email and how it is archived

- They want a system that is easy to use

- Range of understanding of the term "archiving"

- Considerable education would be needed on records management purposes, as distinct from personal archiving

Design – Risk Analysis

- Volume of stored email hard to manage
- Inability to produce vital email records required:
  - Finding "the wood for the trees"
  - Searching across multiple mailboxes
  - Email lost when staff leave the University
  - Not available through staff illness/absence
  - Deleted by staff accidentally or on purpose
  - Stored off the server (e.g. POP, .PST files)

Design – Risk Analysis (continued)

- Proving authenticity

- Inappropriate email being kept and discovered

- Confidential or personal email being disclosed inappropriately

- Email being held longer than it should be

- Risk of losing email records when systems change in the future

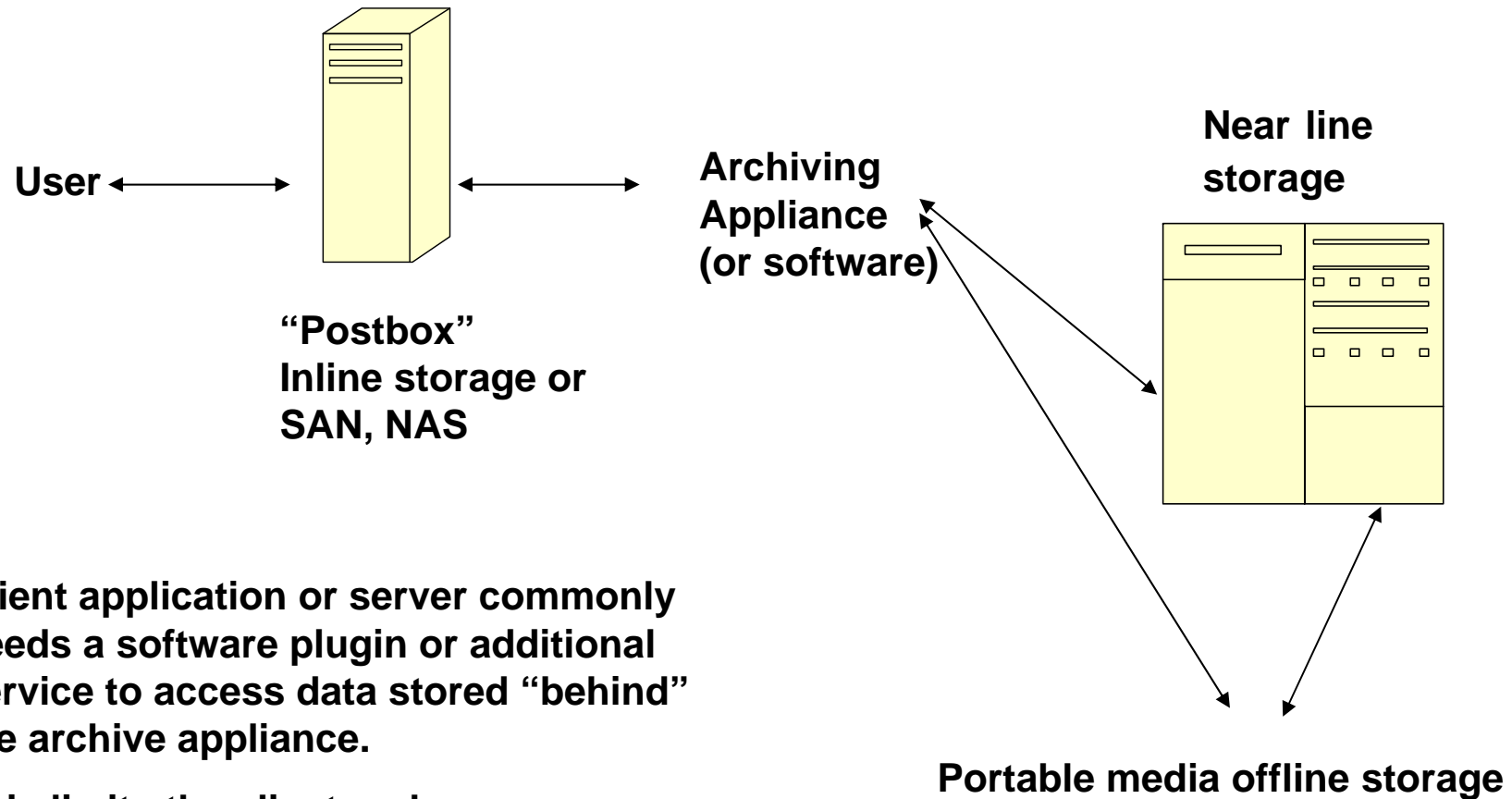- Lost opportunities for future historical research

Design Considerations - Cost

- Software licences

- Archiving appliance cost

- Near line disk storage – Loughborough currently accepts 400000 emails a week (we reject 1.2m per week!). If all of these are stored with attachments, approx 5TB of disk space needed per year at current figures

- Offline disk storage (tape/cd/WORM)

Technical Design - Integration

- Many proprietary archiving solutions are heavily tied into specific email systems and mass storage solutions.

- Many only support Exchange, Notes, Groupwise etc. when using Outlook as client

- A (very) few are agnostic and work with RFC standards

- If you already have a mass storage device, you may be forced to use one archive system and one email system

Design – "Common" Architecture

User ←——————→ [Server] ←——————→ **Archiving Appliance (or software)**

Near line storage

"Postbox" Inline storage or SAN, NAS

Portable media offline storage

Client application or server commonly needs a software plugin or additional service to access data stored "behind" the archive appliance.

This limits the client and server platforms to those specified by the appliance or software vendor.

Technical Design – Catering for diverse user base

- Many UK universities have a diverse range of clients using different operating systems.

- Mac and Linux users may find their options extremely limited as to how they can interact with an archiving system

- Many users don't want their email client prescribed, but many archive systems assume Outlook is the client.

- Non-Outlook users may have to use a separate interface to the archive rather than their preferred email client

Loughborough Implementation

- Implementing archiving as part of a change of email server software and hardware platform

- Decision not to purchase an archiving solution.

- Archive email as "normal email", because:
  - Aids searching
  - Uses familiar interface – normal email client
  - Works across user choice of platforms
  - No specialist application to manage …
  - … or buy

- Disadvantage – storing multiple copies of uncompressed email

Loughborough Implementation – Technical 1

- Accessible via all available client methods
- No additional client application required
  - No training other than user familiarisation
  - Consistent interface
  - Users need no knowledge of storage, just use folders
- Web interface for administrators to control access, retention periods
  - Consistent with folder views in clients
  - Well received by end users due to simplicity
- Easily developed with server platform (CommuniGate Pro) via vendor's published API (Perl module), IMAP protocol, CGI with Apache

Loughborough Implementation – Technical 2

- 2 tier storage – "live", archive (and others)
- Quota applied to primary email storage (limited by volume) on tier 1
- 250Mb initial allowance
- Users can increase to 350Mb via Web page
- Further increases on request where needed – e.g. regular exchange of large files
- Plus "Two Year Store" (limited by time) on tier 2 – more generous quota
- Migrated mail on tier 2 – retained for 2 years
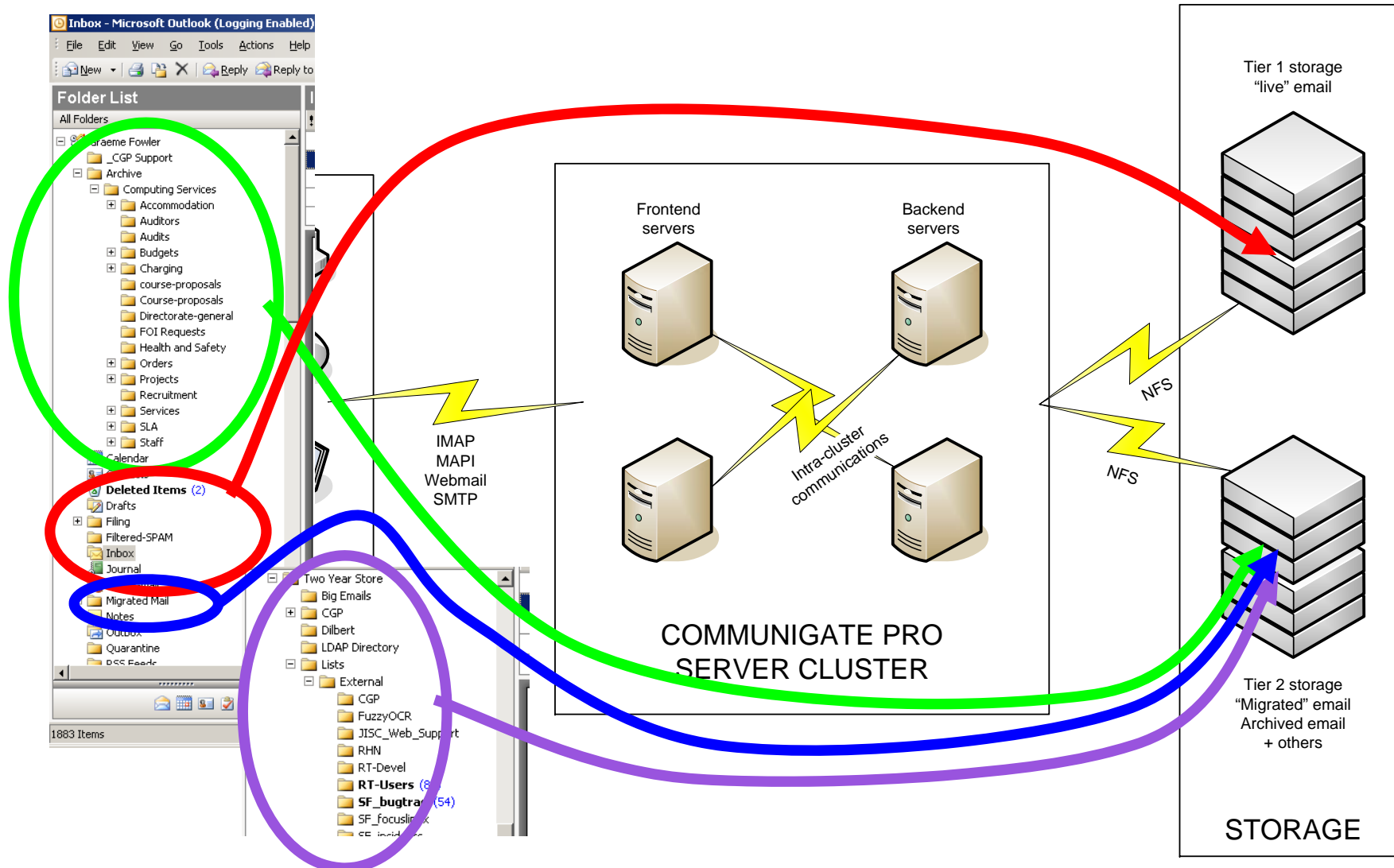
Loughborough Implementation – Policy

- Each department defines its own hierarchical folder structure

- Stored on Tier 2 storage – no quota applied

- Method for selective archiving - "proper records management"– simple drag and drop into folders in predefined structure.

- End user training as part of new email rollout

So how is it going?

- Structure, shared archives are really taking off in some areas
  - 96000 items in 3700 folders so far
  - 22 depts
- Difficulties where archive structure hasn't been put in place before rollout, because...
- Having a quota is a surprise (or a shock!)
- "Filers" like to be involved in structural definition
- "Non-filers" have a learning curve.

Questions

(Garry could, or would, answer them but he's not here)

Questions?

???