Building preservation into the workflow
*The RepoMMan and REMAP projects*

Chris Awre, Richard Green and Simon Lamb
PASIG
Baltimore, November 2008
The University of Hull

Founded 1929

20,000 students (UG and PG)

1,000 faculty

Multidisciplinary (Faculties cover Sciences, Engineering, Arts, Social Sciences, Education, Law, Medicine, Nursing, Business)

Library and archives hold over 1,000,000 items

Mid-sized for the UK
View of a repository

Repositories are often
- Centrally managed
- Geared towards completed materials
- Focused on the collections, not the content creators or owners

This is good where existing digital content requires management and/or preservation

The approach can lead to challenges where content exists but the need for its management/preservation is not as clearly recognised
- Lack of submission
- Lack of provision for materials in development
- A perception by content creators that repositories are distant
The University of Hull repository

All digital content is important

- There is a need for mechanisms to meet differing and flexible needs, throughout the lifecycle of the content
- There is also a need to manage many different types of digital content, and relationships between them
  - Working across institutional and content silos

University of Hull requirements

- Scaleable
  - Digital content is only going to grow
- Standards-based (open standards where possible)
  - To be able to integrate with other institutional systems
  - To provide a future-proof exit strategy
The repository vision
RepoMMan project

Repository Metadata and Management

- JISC-funded project, 2005-7
- Part of the JISC Digital Repositories Programme

Two strands

- User requirements
  - What do our users need a repository for?
- Technical
  - Develop a BPEL and Web Services based workflow tool, for Fedora
  - Investigate and test the automated generation of metadata

These strands seek to achieve and enable personalised information management through a personal repository space
BPEL and Web Services

Business Process Execution Language

- OASIS standard, currently at version 2.0
- A language for specifying business process behaviour based on Web Services
- A mechanism for orchestrating Web Service interactions between systems
- BPEL can draw on any available Web Service, local or external

BPEL has been used to coordinate interaction between Fedora and other institutional systems via Fedora’s Web Service interfaces

- API-M, API-A
Three-tier stack

- LDAP / IdMS
- REPOSITORY SERVER
  - Fedora
  - Fedora web svcs
  - Image Magick
  - MIME Magic
  - Data Fountains
  - JHOVE
  - BuildObject
  - Base64Encode
- BPEL
- University SAN

Confirm deposit to user
Embedding repository interaction

Need to map to existing activity

• File management and FTP paradigms

Basic workflows

• Put object in repository (within identifiable structure)
• Get object from repository
• Delete
• Add metadata
• Share
• Publish
  - Move from personal repository to public one
The personal repository interface
Moving the object through its lifecycle

Private repository assists with the management of digital objects during creation

Once ready for wider access, the object can

- have metadata generated and attached
- be shared with colleagues
  - This is reliant on groups management capability
- be published into the public repository
  - This doesn’t mean it is available on ‘open’ access, but that it resides in the public repository

The publishing step is another BPEL process

- Additional functionality can be incorporated
Lifecycle
REcords MAnagement and Preservation project

- JISC-funded project, 2007-9
- Part of the Repositories and Preservation Programme

Aims

- To identify how an institutional digital repository can support records management and digital preservation (RMDP)
  - Noting that RM is a first stage in DP for many items
- To extend the RepoMMan workflow model to enable workflows that support the full lifecycle of digital objects
  - From creation to withdrawal or perpetual preservation
REMAP overview

REMAP allows us to embed time-driven workflows into an object

Three categories of RMDP alert have been identified:

- **Events 1**: FYI, this has just happened
  - The item you submitted for publishing is now available at http://.....

- **Events 2**: This has happened, that needs to be done
  - An item has been deposited in the accession queue and needs attention

- **Dates 1**: It is now xx days since you requested deposit of... or The deadline of dd/mm/yy has passed

- **Dates 2**: Specified lifespan reached. Hide?

- **Status**: The repository contains nn objects of filetype zzz
REMAP Methodology

As objects are submitted to a publishing queue each will have an ‘flags’ datastream added

Flags datastream written in iCAL (xCAL)

Messages/alerts etc. managed by Darwin Calendar Server

Also looking at Bedeworks

Flags datastream in the Fedora object means that the system can be reconstructed in the event of a crash
REMAP use cases

Range of use cases investigated to date

- Preparation of committee papers by Committee Section
- Past examination papers as learning resources
- Learning and Teaching Programme approvals
- University Register of Policies and Procedures
- University Quality Handbook
- The RepoMMan publication process
- Support for Electronic Theses and Dissertations (ETD)
- Digital Archives at the University of Hull
- RMDP for Spoken Word Services (large audio-visual archive)
Register of policies and procedures

1. Policy or procedure deposited into the Register – at any time of the year
   - Published on portal
2. Annual review in May
   - Annual review of policies and procedures is carried out
3. Authors of policies are contacted by email and asked to review the policies for accuracy and relevance
4. Details on Register are amended
   - Date revised on document
   - Possibly Date for review – depends on the time of year document was amended e.g. if amended in March wouldn’t be in the annual review for May
5. Document is amended
   - Date for review is amended on Register
6. Document is checked but remains unchanged
   - Date for review is amended on Register
7. Document is checked and is no longer relevant
   - Document is deleted

Notes:
- previous versions of policies and procedures are not currently kept – this is a potential risk
- Response to the review is usually 70-80% which mean that all the policies and procedures are not officially checked on annual basis
Multimedia management
Proactive preservation management

Some of these lifecycle workflows may have or require preservation stages

Preservation web services can be linked into BPEL workflows to carry out these stages, e.g.,

- PRONOM/DROID
- JHOVE
- CRiB
- Planets
- SHERPA DP2 / SOAPI

BPEL allows services to be interchanged according to availability and policy
Institutional preservation issues

To install services locally or link to external services?

To install locally

- Control over service delivery
- Resource required to maintain service

To link to external services

- Benefit from network level services
- Possible issues around service reliability

Hull currently favours the latter approach, using BPEL and web services as the framework to develop preservation workflows
Thank you

Chris Awre
- c.awre@hull.ac.uk

Richard Green
- r.green@hull.ac.uk

University of Hull edocs repository
- http://edocs.hull.ac.uk

Acknowledgements
- Simon Lamb, Gary Thompson, Robert Sherratt, Vicky Mays, Judy Burg
- Spoken Word team, Glasgow Caledonian University
- Fedora Commons

Projects
- RepoMMan - http://www.hull.ac.uk/esig/repomman/
- REMAP - http://www.hull.ac.uk/remap/