

Enabling digital preservation through an institutional repository *The REMAP project*

Chris Awre, Richard Green and Simon Lamb
Digital Preservation Roadshows 2009
York, 26th July 2009

View of an institutional digital 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 and this is acknowledged

This 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

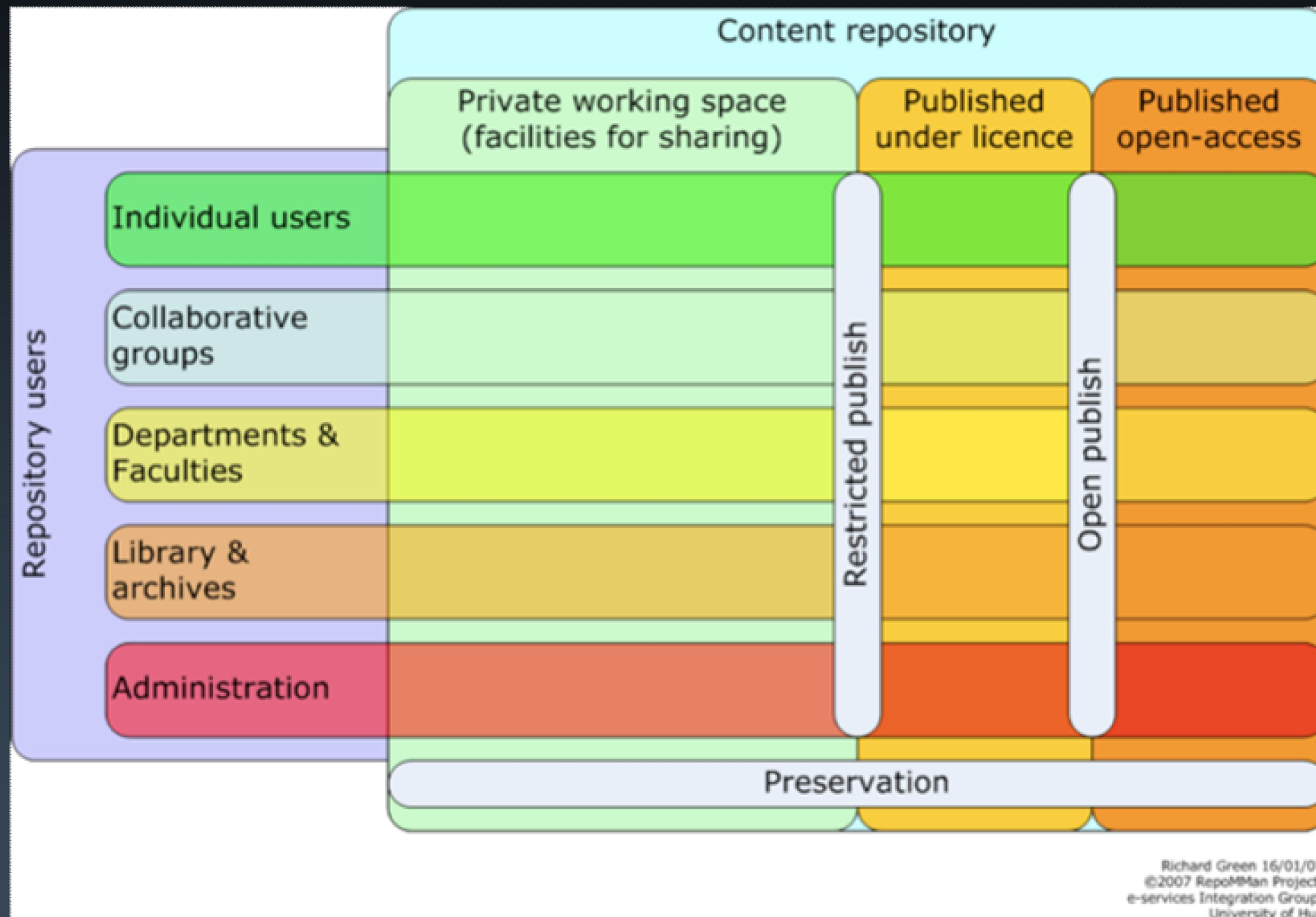
All digital content is important

- 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

- Scalable
 - 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



Fedora digital repository

A powerful community source repository system

- Development has been overseen by Fedora Commons, a not-for-profit foundation, launched September 2007
 - Fedora Commons has now merged with DSpace to form DuraSpace
- Fedora 3.2 released May 2009

Organisation

- Core development team within the Commons
- Community input guides development
 - Solutions communities being developed
- Collaboration with DSpace to work on common goals and highlight differences

Features of Fedora

Powerful digital object model

Extensible metadata management

Expressive inter-object relationships

Web service integration

Version management

Configurable security architecture

OAI-PMH conformance

Preservation worthy

Features of Fedora that support digital preservation

Powerful digital object model

Extensible metadata management

Expressive inter-object relationships

Web service integration

Version management

Configurable security architecture

OAI-PMH conformance

Preservation worthy

Fedora started with a focus on

ORGANISATION

This has matured and evolved into an emphasis on

DURABILITY

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

REMAP project



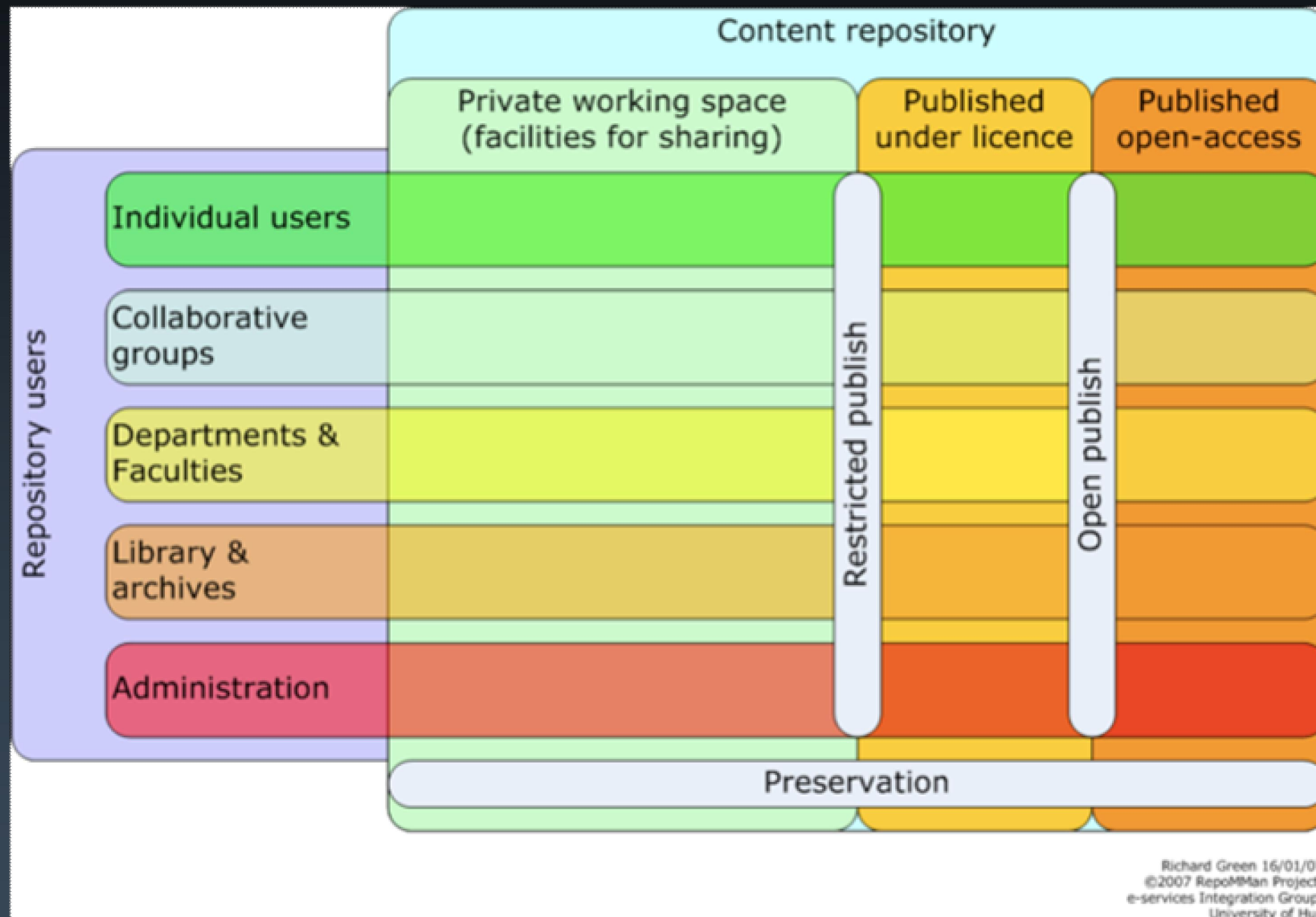
REcords MAnagement and Preservation project

- 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
 - Includes publishing from personal to public repository

The repository vision



REMAP overview

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

Three categories of RMDP alert have been identified:

- Events
 - This has happened, that needs to be done, e.g., an item has been deposited in the accession queue and needs attention
- Dates
 - Specified lifespan reached. Hide?
- Status
 - The repository contains nn objects of filetype zzz

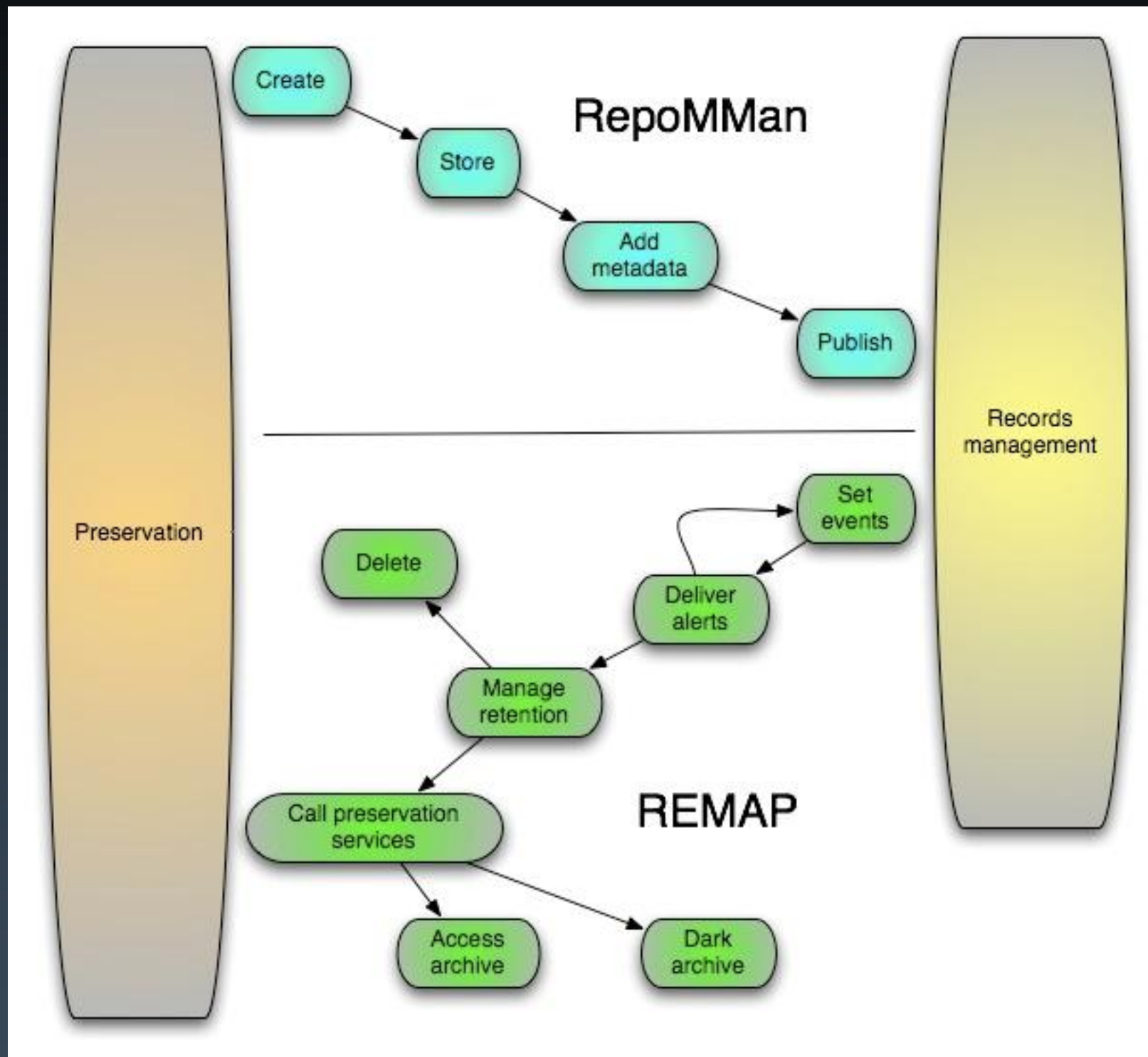
13

REMAP use cases

Range of use cases investigated to date

- Preparation of committee papers
- Past examination papers
- University Register of Policies and Procedures
- University Quality Handbook
- Electronic Theses and Dissertations (ETD)
- Digital Archives at the University of Hull
- RMDP for Spoken Word Services (large audio-visual archive at Glasgow Caledonian University)

Lifecycle



REMAP Methodology

As objects are submitted to a publishing queue each can have an RMDP alert added

- RMDP alert written in iCAL
- Alerts are managed by a calendar server

The publishing process can also call on available Web Services to carry out preservation processes

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

DROID / PRONOM / JHOVE

DROID is installed as a local service and ‘wrapped’ as a Web Service so we can include it in workflows

Objects are passed through DROID as part of the publishing process

This produces a file signature, which is then passed to PRONOM via Web Services

- The resultant file format information received is stored alongside the object for future reference

JHOVE is used to capture image information

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

Using the repository

University Repository



Upload Manager

 **Browse**

 **Upload**



Repository

- |  | REMAP materials |
|---|-----------------------------|
|  | ClinPsyD-08-pmi-babiker.pdf |
|  | hydra-page-block.jpg |
|  | OR08F-RMDP-v10.doc |
|  | OR08F-RMDP-v10.pdf |

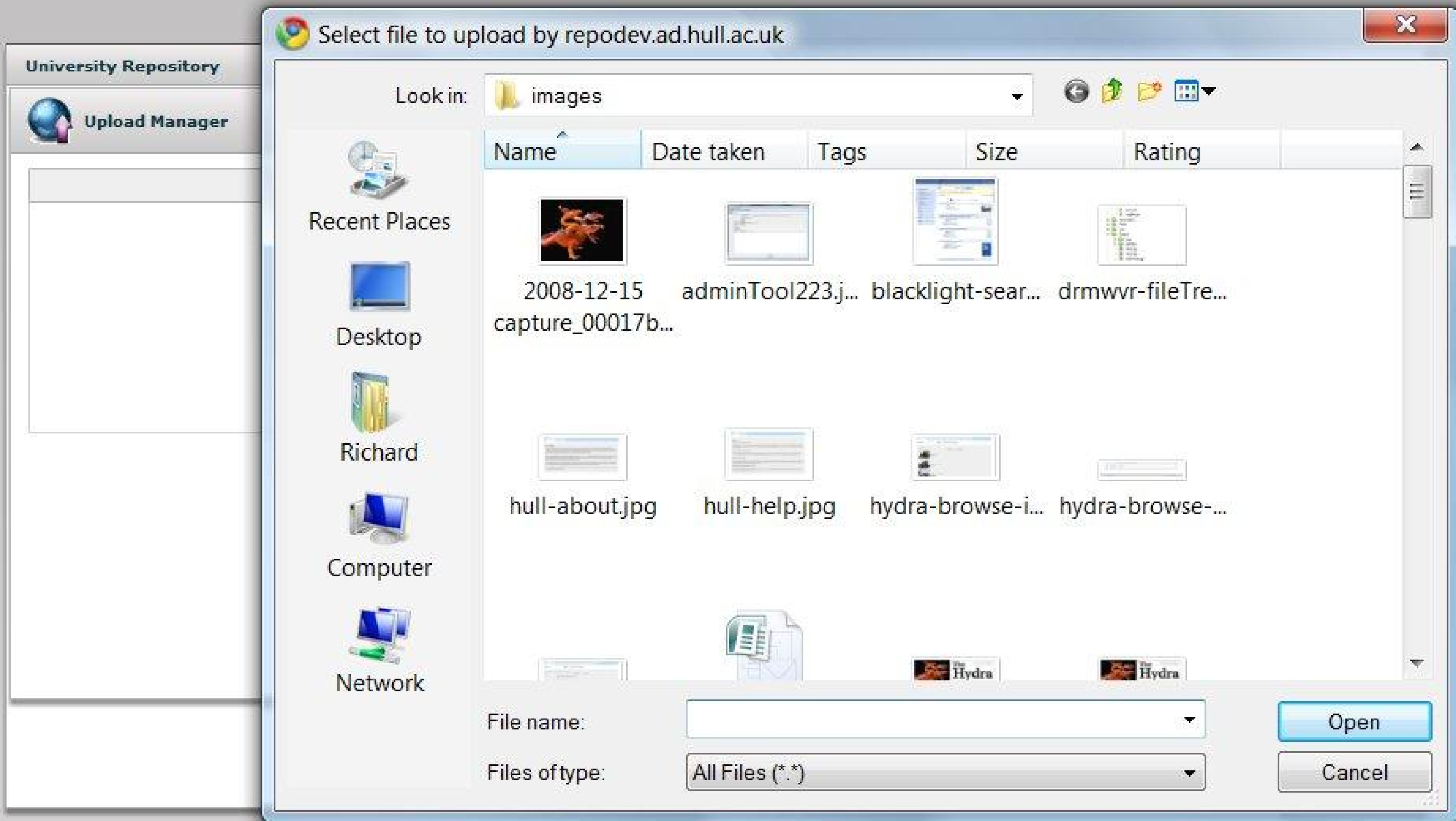
 **Download**

 **Create Folder**

 **Delete**

 **Publish**

Depositing



File system paradigm

University Repository




Upload Manager

hydra-page-block.jpg

 Browse

Progress 0%

 Upload



Repository



REMAP materials



ClinPsyD-08-pmi-babiker.pdf



hydra-page-block.jpg



OR08F-RMDP-v10.doc



OR08F-RMDP-v10.pdf



Download



Create Folder



Delete



Publish

Versioning

The screenshot displays a web interface for a University Repository. On the left is the 'Upload Manager' section, and on the right is the 'Repository' section. The 'Repository' section lists several files, with 'hydra-page-block.jpg' selected. An 'Advanced' modal window is open, showing details for the selected file. The modal includes a file icon, the filename 'hydra-page-block.jpg', and a PID field containing 'hullwf:21'. Below this, a table lists two versions of the file, both dated '23 February 2009'. The first version is at '14:17:31' and the second at '14:16:53'. The second version is highlighted in blue. To the right of the table are buttons for 'Publish', 'Share', 'Delete', and 'Download'. At the bottom right of the modal is a 'Close' button.

University Repository

Upload Manager

Repository

REMAP materials

ClinPsyD-08-pmi-babiker.pdf

DORSOL2-proposal.doc

hydra-page-block.jpg

OR08_abstract.pdf

Advanced

hydra-page-block.jpg

PID: hullwf:21

23 February 2009	14:17:31
23 February 2009	14:16:53

Publish

Share

Delete

Download

Close

Publishing

University Repository

Upload Manager

Repository

Publish

What Type of Object do you want to Publish?

- Software
- Sound
- Other
- ▼ Text
 - Book
 - Book Chapter
 - Conference paper or abstract
 - Examination paper
 - Handbook

DC

Cancel

Next

Publish

Metadata

University Repository

Upload Manager

Publish

Ownership DC

Creator : Green, Richard A

Publisher : The University of Hull

Rights :

- ☐ © 2009 The University of Hull
- ☐ © The author. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright holder.
- ☐ Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: <http://creativecommons.org/licenses/by-nc-sa/2.0/uk/>

Back **Cancel** **Next**

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/>