Chris Awre

UKCoRR meeting
Teeside University
9th November 2012
An institutional repository

One institution = one repository?

• Repositories are infrastructure
  – Maintaining infrastructure requires resource, which we need to minimise to justify costs in the long-term

• Content doesn’t sit in silos
  – One repository facilitates cross-fertilisation of use

• Integration with one system
  – Embedding the repository means linking to one place
Five principles

- A repository should be content agnostic
- A repository should be (open) standards-based
- A repository should be scalable
- A repository should understand how pieces of content relate to each other
- A repository should be manageable with limited resource
Five principles (leading to our implementation)

- Fedora is content agnostic
- Fedora is (open) standards-based
- Fedora is scalable
- Fedora understands how pieces of content relate to each other
- Fedora is manageable with limited resource
  — With help from the community
Hydra

• A collaborative project between:
  – University of Hull
  – University of Virginia
  – Stanford University
  – Fedora Commons/DuraSpace
  – MediaShelf LLC

• Unfunded (in itself)
  – Activity based on identification of a common need

• Aim to work towards a reusable framework for multipurpose, multifunction, multi-institutional repository-enabled solutions

• Timeframe - 2008-11 (but now extended indefinitely)
Fundamental Assumption #1

No single system can provide the full range of repository-based solutions for a given institution’s needs,

...yet sustainable solutions require a common repository infrastructure.

Fundamental Assumption #2

No single institution can resource the development of a full range of solutions on its own,

...yet each needs the flexibility to tailor solutions to local demands and workflows.
Fedora and Hydra

• Fedora can be complex in enabling its flexibility

• How can the richness of the Fedora system be enabled through simpler interfaces and interactions?
  – The Hydra project has endeavoured to address this, and has done so successfully
  – Not a turnkey, out of the box, solution, but a toolkit that enables powerful use of Fedora’s capabilities through lightweight tools
    • Principles can also be applied to other repository environments

• Hydra ‘heads’
  – Single body of content, many points of access into it
Hydra is a Repository Solution

Hydra is a repository solution that is being used by institutions on both sides of the North Atlantic to provide access to their digital content. Hydra provides a versatile and feature-rich environment for end-users and repository administrators alike.

Hydra is a Community

Hydra is a large, multi-institutional collaboration. The project gives like-minded institutions a mechanism to combine their individual repository development efforts into a collective solution with breadth and depth that exceeds the capacity of any single institution to create, maintain or enhance on its own. The motto of the project’s partners is “if you want to go fast, go alone. If you want to go far, go together.”

Hydra is a Technical Framework

Hydra is an ecosystem of components that lets institutions deploy robust and durable digital repositories (the body) supporting multiple “heads”: fully-featured digital asset management applications and tailored workflows. Its principle platforms are the Fedora Commons repository software, Soir, Ruby on Rails and Blacklight. See how you can get started.

Hydra is Open Source Software

Hydra software is free and open source, available under an Apache 2 license.
Hydra partnerships

• From the beginning key aims have been and are:
  – to enable others to join the partnership as and when they wished
    (Now up to 10 partners, with two others in process)
  – to establish a framework for sustaining a Hydra community as much
    as any technical outputs that emerge
    • Establishing a semi-legal basis for contribution and partnership

“If you want to go fast, go alone. If you want to go far, go together”

(African proverb)
Community Model

Hydra Steering Group
- small coordinating body
- collaborative roadmapping (tech & community)
- resource coordination
- governance of the "tech core" and Hydra Framework
- community mtce. & growth

Currently
- DuraSpace
- Hull
- MediaShelf
- Stanford
- Virginia

Hydra Partners
- shape and direct work
- commission "Heads"
- functional requirements & specs
- UI design & spec
- Documentation
- Training
- Data & content models
- "User groups"

Hydra Developers
- define tech architecture
- code development
- integration & release

Founders
- Duraspace
- Hull
- Stanford
- UVa

Committers
Contributors
Tech. Users
Hydra technical implementation

• Fedora
  – All Hydra partners are Fedora users

• Solr
  – Very powerful indexing tool, as used by...

• Blacklight
  – Prior development at Virginia (and now Stanford/JHU) for OPAC
  – Adaptable to repository content

• Ruby
  – Agile development / excellent MVC / good testing tools

• Ruby gems
  – ActiveFedora, Opinionated Metadata, Solrizer (MediaShelf contributions)
Four Key Capabilities

1. Support for any kind of record or metadata
2. Object-specific behaviors
   - Books, Articles, Images, Music, Video, Manuscripts, etc.
3. Tailored views for domain or discipline-specific materials
4. Easy to augment & over-ride with local modifications
## Work of faculty and students

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Disseminate research outputs / learning materials</td>
<td>- Disseminate theses / dissertations</td>
</tr>
<tr>
<td>- Manage research data</td>
<td>- Provide exam papers</td>
</tr>
<tr>
<td>- Archive event outputs</td>
<td>- Student handbook archive</td>
</tr>
</tbody>
</table>

Granular security required to manage these different activities

The repository has been tied into our institutional CAS system

Materials can be open, internal, or restricted to user groups / users
Records of events and performance

Creative writing – discussions with authors

Inaugural lectures

University Learning & Teaching Conference

Campus-based e-publishing

Integration with Open Journal Systems to enable archiving of publications
Experimental and observational data

- Tiptoeing into data management

- JISC History DMP project
  - Identified ways to encourage and facilitate the planning of data management

- EPSRC roadmap
  - Highlighting ways forward to make the most of the data we produce
Adapt to the content

Journal article

CLIF : moving repositories upstream in the content lifecycle

Authors
Waddington, Simon; Green, Richard A.; Awe, Christopher L.

Subjects
CLIF, JISC, Content lifecycle, Institutional repository, Sakai, eBridge, Microsoft SharePoint

Rights
Creative Commons Licence: Attribution-NonCommercial-Share Alike 2.0 UK: England and Wales. See: http://creativecommons.org/licenses/by-nc-sa/2.0/uk/

Abstract
The UK JISC-funded Content Lifecycle Integration Framework (CLIF) project has explored the management of digital content throughout its lifecycle from creation through to preservation or disposal. Whilst many individual systems offer the capability of carrying out lifecycle stages to varying degrees, CLIF recognised that only by facilitating the movement of content between systems could the full lifecycle take advantage of systems specifically geared towards different stages of the digital lifecycle. The project has also placed the digital repository at the heart of this movement and has explored this through carrying out integrations between Fedora and Sakai, and Fedora and SharePoint. This article will describe these integrations in the context of lifecycle management and highlight the issues discovered in enabling the smooth movement of content as required.

Date
2012

Language
English

Publishers
The University of Hull, Texas Digital Library

Published

Published in
Journal of digital information, 2012

ISSN (Electronic)
1368-7506

Publisher
Texas Digital Library

Volume
13

Issue
1

Dataset

HMAP Dataset 06: Newfoundland, 1675-1698

Person
Pope, P. (Author)

Subjects
Population census, History of marine animal populations, Fishing effort, Cod fishery

Rights
Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: http://creativecommons.org/licenses/by-nc-sa/2.0/uk/

Citations
(a) The dataset, please cite as follows: P. Pope, ed., 'Newfoundland, 1675-1698' in M.G Barnard and J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap). (b) Supporting documentation: please cite as follows: P. Pope, ‘HMAP dataset 6: Newfoundland, 1675-1698, Supporting Documentation’, in M.G Barnard and J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap)

Description
Fishermen, settlers and cod catches in 17th-century Newfoundland.

The map below gives an indication of the extent of the Newfoundland-Labrador shelf; the 'view as map' link in the download panel at the right will show a much more detailed representation. The kml file download, when used with Google Earth, will render the extent of the Newfoundland-Labrador shelf in detail.

Coverage
Newfoundland

Temporal
1675-1698

Geo-data

Downloads

- Database - ASCII format
- Database - Access 2000 format
- Database - csv format
- Documentation - PDF format
- Documentation - text format
- Newfoundland-Labrador Shelf - kml file

View as map

UKCoRR – Hydra | 9 November 2012 | 16
Organise the content

Processed HMAP Datasets
---Law
---Scarborough School of Education
---Campus Based Publishing System (CBPP)
---cbpplatform objects
---hull-cbpp objects
---Committee Section
---General documents
---Datasets
---Domesday
---Databases
---Documentation
---Phillimore appendices
---Phillimore county introductions
---Phillimore county notes
---Statistics notes
---Translation
---English
---Repertorium of Middle English Prose Sermons
---History of Marine Animal Populations (HMAP)
---Processed HMAP Datasets

Structural sets

Larkin Centre events
Resource type: Display set
Description: A collection of recordings made at public events organised by the University’s Larkin Centre featuring interviews with modern writers.

History of Marine Animal Populations (HMAP) datasets
Resource type: Display set
Description: The HMAP datasets are a research resource comprising of information derived largely from historical records relating to fishing catches and effort...

Skills for scientists: employability resource pack
Resource type: Display set
Description: This pack is a structured resource that can be used as a complete module or as individual exercises to support other career-based activities. The a...

Domesday dataset
Resource type: Display set
Subject: Domesday Book
Description: A collection of data about and around the 1086 Domesday Book. The dataset was originally lodged in the repository by Professor John Palmer in 2008....
Digital archives management

- Archives colleagues took part in Mellon-funded AIMS project
  - An Inter-institutional Model for Stewardship of born-digital collections

  Developing practice for archivists in dealing with born-digital collections through events and advocacy

  Using Hydra to develop tools that help put the model into practice
Finch outcomes

• Whatever you may think of the Finch Report (Government report on expanding access to research publications)
  – It did highlight the valuable role repositories can play in providing supporting infrastructure

• It focused on the need to have a way of managing:
  – Different types of grey literature
  – Theses/dissertations
  – Links between research data and publications
  – Preservation

• Looks like an opportunity!
Onwards...

Our repository is still a work in progress - probably always will be

...it is maturing

We have been able to apply it to a range of purposes

It is a repository for the institution as a whole – Hydra has helped enable this
Thank you

Chris Awre – c.awre@hull.ac.uk

Simon Lamb – s.lamb@hull.ac.uk

Richard Green – r.green@hull.ac.uk

Hydra at Hull – http://hydra.hull.ac.uk

Hydra Project – http://projecthydra.org

Hydra UK event, LSE, 22nd November 2012